

**Semismooth Newton methods for
operator equations in function spaces**

Michael Ulbrich

April 2000

TR00-11

Department of Computational and Applied Mathematics - MS134
Rice University
6100 Main Street
Houston, Texas 77005-1892
USA

SEMISMOOTH NEWTON METHODS FOR OPERATOR EQUATIONS IN FUNCTION SPACES

MICHAEL ULBRICH*

Abstract. We develop a semismoothness concept for nonsmooth superposition operators in function spaces. The considered class of operators includes NCP-function-based reformulations of infinite-dimensional nonlinear complementarity problems, and thus covers a very comprehensive class of applications. Our results generalize semismoothness and α -order semismoothness from finite-dimensional spaces to a Banach space setting. Hereby, a new generalized differential is used that can be seen as an extension of Qi's finite-dimensional C-subdifferential to our infinite-dimensional framework. We apply these semismoothness results to develop a Newton-like method for nonsmooth operator equations and prove its local q -superlinear convergence to regular solutions. If the underlying operator is α -order semismoothness, convergence of q -order $1 + \alpha$ is proved. We also establish the semismoothness of composite operators and develop corresponding chain rules. The developed theory is accompanied by illustrating examples and by applications to nonlinear complementarity problems.

Key words. Newton-like methods, semismoothness, superposition operators, generalized differentials, nonlinear complementarity problems, superlinear convergence

AMS subject classifications. 49M15, 65K05, 90C33, 49J52, 47J25, 47H30

1. Introduction. The notion of semismoothness was introduced by Mifflin [34] for real-valued functions defined on finite-dimensional spaces, and extended to mappings between finite-dimensional spaces by Qi [40] and Qi and Sun [42]. The importance of semismooth equations results from the fact that, although the underlying mapping is in general nonsmooth, Newton's method is still applicable and converges locally with q -superlinear rate to a regular solution [39, 40, 42]. For related early approaches to nonsmooth Newton methods we refer to [32, 33, 37].

Written in a form most convenient for our purposes, a mapping $f : \mathbb{R}^k \rightarrow \mathbb{R}^l$ is called semismooth at x if f is Lipschitz near x , directionally differentiable at x , and if

$$(1.1) \quad \max_{M \in \partial f(x+h)} \|f(x+h) - f(x) - Mh\| = o(\|h\|) \quad \text{as } h \rightarrow 0,$$

where ∂f denotes Clarke's generalized Jacobian [11]. Further, f is called α -order semismooth, $1 < \alpha \leq 1$, if the order in (1.1) can be improved to $O(\|h\|^{1+\alpha})$.

The objective of this paper is to extend the notions of semismoothness and α -order semismoothness, respectively, to nonlinear superposition operators in function spaces, and to develop a corresponding superlinearly convergent Newton-like method. Motivated by applications arising in mathematical modeling and optimal control, we consider operators of the form

$$(1.2) \quad \Psi : Y \rightarrow L^r(\Omega), \quad \Psi(y)(\omega) = \psi(F(y)(\omega)),$$

with mappings $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$ and $F : Y \rightarrow \prod_{i=1}^m L^{r_i}(\Omega)$, where $1 \leq r \leq r_i < \infty$, Y is a real Banach space, and $\Omega \subset \mathbb{R}^n$ is a bounded open domain. Essentially, our working assumptions are that ψ is Lipschitz continuous and semismooth, and that F is continuously Fréchet-differentiable. The detailed assumptions are given below. The main result of this paper is a semismoothness-like estimate of the form

$$(1.3) \quad \sup_{M \in \partial_s \Psi(y+s)} \|\Psi(y+s) - \Psi(y) - Ms\|_{L^r} = o(\|s\|_Y) \quad \text{as } s \rightarrow 0 \text{ in } Y.$$

*Zentrum Mathematik, Technische Universität München, D-80290 München, Germany (mulbrich@ma.tum.de). The author was supported by Deutsche Forschungsgemeinschaft (DFG) grant U1157/3-1 and by CRPC grant CCR-9120008.

We also give conditions under which the remainder term in (1.3) is of the order $O(\|s\|_Y^{1+\alpha})$, $0 < \alpha \leq 1$. In this case we call Ψ α -order semismooth. The multifunction (i.e., set-valued mapping) $\partial_s \Psi : Y \rightrightarrows \mathcal{L}(Y, L^r)$ denotes an appropriate vector-valued generalized differential of Ψ , which can be interpreted as a generalization of Qi's finite-dimensional C-subdifferential [41] to operators in function spaces. The estimate (1.3) generalizes (1.1) to the function space setting. We will not require that Ψ be directionally differentiable, because this is not needed in the analysis of Newton's method.

Based on (1.3), we develop a locally q -superlinearly convergent Newton method for the nonsmooth operator equation

$$(1.4) \quad \Psi(y) = 0.$$

Moreover, in the case where Ψ is α -order semismooth we prove convergence with q -rate $1 + \alpha$. In analogy to BD-regularity assumptions for finite-dimensional semismooth Newton methods, we impose a regularity condition on the elements of the generalized differential. Further, as was already observed earlier in the context of related local convergence analyses in function space [30, 52], we have to incorporate a smoothing step to overcome the non-equivalence of norms. We also will provide an example showing that this smoothing step is indispensable.

Recently, a different semismoothness concept for operator equations was proposed by Chen, Nashed and Qi [10]. We point out that our approach differs significantly from the one in [10]. There, the notion of a slanting function is introduced and a generalized derivative, the slant derivative, is obtained as the collection of all limits of the slanting function as $y_k \rightarrow y$. Semismoothness is then defined by imposing appropriate conditions on the approximation properties of the slanting function and the slant derivative.

Although the differentiability properties of superposition operators with smooth ψ are well investigated, see, e.g., the expositions [6] and [7], this is not the case for nonsmooth functions ψ . Further, even if ψ is smooth, for operator equations of the form (1.4) the availability of local convergence results for Newton-like methods appears to be very limited.

As an important application of our results, which originally motivated our investigations, we discuss the reformulation of the nonlinear complementarity problem (NCP) posed in function space as an equivalent nonsmooth operator equation of the form (1.4). This problem consists in finding $y \in L^p(\Omega)$ such that almost everywhere on Ω holds

$$(1.5) \quad y \geq 0, \quad Z(y) \geq 0, \quad yZ(y) = 0,$$

where the operator $Z : L^p(\Omega) \rightarrow L^r(\Omega)$, $1 \leq r < p \leq \infty$, is given. In order to reformulate (1.5) as a nonsmooth operator equation, we use an NCP-function [47], i.e., a function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ satisfying

$$\phi(x) = 0 \quad \iff \quad x_1 \geq 0, \quad x_2 \geq 0, \quad x_1 x_2 = 0,$$

to rewrite the pointwise complementarity conditions in (1.5) as equations. Doing this, (1.5) can be written equivalently in form of the operator equation

$$(1.6) \quad \Phi(y) = 0, \quad \text{where} \quad \Phi(y)(\omega) \stackrel{\text{def}}{=} \phi(y(\omega), Z(y)(\omega)), \quad \omega \in \Omega.$$

Obviously, choosing $Y = L^p(\Omega)$, $r_1 = r_2 = r$, and $F : y \in Y \mapsto (y, Z(y))$, we have $\Psi \equiv \Phi$ with Ψ as in (1.2). For finite-dimensional NCPs, this approach of reformulating the problem as an equivalent system of equations has been extensively studied in recent years [14, 16, 17, 28, 29, 50] and led to very efficient Newton-like methods, see, e.g., [35]. To avoid imposing

a strict complementarity condition, it is advantageous to work with nondifferentiable rather than smooth NCP-functions. Since the introduction of the semismooth Fischer–Burmeister NCP-function

$$(1.7) \quad \phi_{FB}(x) = \sqrt{x_1^2 + x_2^2} - x_1 - x_2$$

by Fischer [19], many researchers agree that semismooth NCP-functions are a very powerful tool to develop efficient algorithms with strong theoretical properties. There are also close connections between the NCP-function approach and non-interior path-following methods for NCPs [9], which recently were introduced and analyzed in finite dimensions. Hereby, the NCP-function ϕ is embedded in a class of smooth perturbations ϕ_σ , where $\sigma \geq 0$ is a parameter. For $\sigma > 0$ the function ϕ_σ is smooth, whereas $\phi_0 = \phi$. For the Fischer-Burmeister function ϕ_{FB} , e.g., the functions ϕ_σ can be obtained by adding the term σ under the square root. The main idea of these methods, transcribed to our setting, consists in following the trajectory of solutions to the corresponding perturbed operator equations $\Phi_\sigma(y) = 0$ as $\sigma \rightarrow 0$. Usually, corrector steps are computed by Newton's method. In the asymptotic phase $\sigma \rightarrow 0$ the behavior of Newton's method on the unperturbed equation plays a key role in achieving fast local convergence. We therefore believe that the results presented in this paper will also be helpful to investigate path-following methods in a function space setting.

We emphasize that the number of applications fitting in our framework is huge, in particular those involving complementarity, see [15, 18, 21, 22, 31, 36, 38]. Many of these applications arise from infinite-dimensional variational inequalities that model systems being continuous in time and/or space [15, 21, 31, 36], and therefore are posed in function spaces. Hence, the development and analysis of efficient abstract algorithms for the solution of the infinite-dimensional problem (1.5) is very desirable in order to derive robust, efficient, and mesh-independent methods for the solution of the discretized problem. The nonsmooth Newton method developed in this paper is directly applicable to NCP-function-based reformulations of the NCP (1.5) and can therefore be seen as a generalization of semismooth Newton methods for finite-dimensional NCPs.

For the purpose of illustration, we briefly describe how a particular optimal control problem can be converted to an NCP of the form (1.5). The problem we describe can be shown to satisfy all assumptions that we are going to impose in this work. For details we refer to [51]. Consider the following distributed optimal control problem of an elliptic partial differential equation with upper bounds on the control:

$$(1.8a) \quad \begin{aligned} & \underset{w \in L^2(\Omega)}{\text{minimize}} && J(w) \stackrel{\text{def}}{=} \frac{1}{2} \|u(w) - u_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2} \|w - w_d\|_{L^2(\Omega)}^2 \\ & \text{subject to} && w \leq \psi \quad \text{on } \Omega, \end{aligned}$$

where $u = u(w) \in H_0^1(\Omega)$ (the usual Sobolev space) is the weak solution of the strongly elliptic state equation

$$(1.8b) \quad - \sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial u}{\partial x_j} \right) = w \quad \text{on } \Omega.$$

We assume $\lambda > 0$, $a_{ij} \in L^\infty(\Omega)$, $u_d \in L^2(\Omega)$, and $w_d, \psi \in L^\infty(\Omega)$. Denoting by $\nabla J(w) \in L^2(\Omega)$ the L^2 -Riesz representation of the gradient of J , it can be shown that \bar{w} solves the control problem if and only if $\bar{y} = \psi - \bar{w}$ solves the NCP (1.5) with $Z(y) = -\nabla J(\psi - y)$. We will further discuss this problem in Example 5.5.

For the development of a semismoothness concept we have to choose an appropriate vector-valued generalized differential for the operator Ψ . Although the available literature on

generalized differentials and subdifferentials is mainly focused on real-valued functions, see, e.g., [8, 11, 12, 45] and the references therein, several authors have proposed and analyzed generalized differentials for nonlinear operators between infinite-dimensional spaces [13, 20, 24, 43, 48]. In our approach, we work with a generalized differential that exploits the structure of Ψ . Roughly speaking, our general guidance hereby is to transcribe, at least formally, componentwise operations in \mathbb{R}^k to pointwise operations in function spaces. To sketch the idea, note that the finite-dimensional analogue of the operator Ψ is the mapping

$$\Psi^f : \mathbb{R}^k \rightarrow \mathbb{R}^l, \quad \Psi_j^f(x) = \psi(F^j(x)), \quad j = 1, \dots, l$$

with ψ as above and C^1 -mappings $F^j : \mathbb{R}^k \rightarrow \mathbb{R}^m$. We have the correspondences $\omega \in \Omega \leftrightarrow j \in \{1, \dots, l\}$, $y \in Y \leftrightarrow x \in \mathbb{R}^k$, and $F(y)(\omega) \leftrightarrow F^j(x)$. Componentwise application of the chain rule for Clarke's generalized gradient [11] shows that the C-subdifferential of Ψ^f consists of matrices $M \in \mathbb{R}^{l \times k}$ having rows of the form

$$M_j = \sum_{i=1}^m d_i^j (F_i^j)'(x), \quad \text{with } d^j \in \partial\psi(F^j(x)).$$

Carrying out the same construction for Ψ in a purely formal manner suggests to choose a generalized differential for Ψ consisting of operators of the form

$$v \in Y \mapsto \sum_{i=1}^m d_i \cdot (F_i'(x)v), \quad \text{with } (d_1, \dots, d_m)(\omega) \in \partial\psi(F(y)(\omega)) \text{ a.e. on } \Omega,$$

where the inclusion on the right is meant in the sense of measurable selections. One advantage of this approach, which motivates our choice of the generalized differential $\partial_s \Psi$, is that it consists of relatively 'concrete' objects as compared to those investigated in, e.g., [13, 20, 24, 43, 48], which necessarily are more abstract since they are not restricted to a particular structure of the underlying operator. It is not the objective of this paper to investigate the connections between the generalized differential $\partial_s \Psi$ and other generalized differentials. There are close relationships, but we leave it as a topic for future research. Here, we concentrate on the development of a semismoothness concept based on $\partial_s \Psi$, a related nonsmooth Newton's method, and the relations to the respective finite-dimensional analogues.

As already mentioned, the literature on Newton-like methods for the solution of nonlinear complementarity problems or, closely related, bound-constrained optimization problems posed in function spaces is very limited. Hereby, we call an iteration Newton-like if each iteration essentially requires the solution of a linear operator equation. We point out that in this sense sequential quadratic programming (SQP) methods for problems involving inequality constraints [1, 2, 3, 4, 5, 23, 49] are not Newton-like, since each iteration requires the solution of a quadratic programming problem (or, put differently, a linearized generalized equation) which is in general significantly more expensive than solving a linear operator equation. Therefore, instead of applying the methods considered in this paper directly to the nonlinear problem, they also could be of interest as subproblem solvers for SQP methods.

Probably the investigations closest related to ours are the analysis of Bertsekas' projected Newton method by Kelley and Sachs [30], and the investigation of affine-scaling interior-point Newton methods by Ulbrich and Ulbrich [52]. Both papers deal with bound-constrained minimization problems in function spaces and establish the local q-superlinear convergence of their respective Newton-like methods. In both approaches the convergence results are obtained by estimating directly the remainder terms appearing in the analysis of the Newton iteration. Hereby, specific properties of the solution are exploited, and a strict complementarity condition is assumed in both papers. We develop our results for the general problem class

(1.4) and derive the applicability to nonlinear complementarity problems as a simple, but important special case. In the context of NCPs and optimization, we do not have to assume any strict complementarity condition. Further, we organize our analysis of Newton's methods by decomposing it in two parts: First, we develop a semismoothness result that replaces differentiability in ordinary Newton methods. Second, an invertibility condition on the members of the generalized differential is introduced. A detailed treatment of the application of the results developed here to optimal control problems and numerical illustrations can be found in the author's paper [51]. In addition, [51] develops a convenient sufficient condition that can be used to verify our regularity condition.

In Section 2 we review some concepts of finite-dimensional nonsmooth analysis that are important in our context, in particular generalized differentials and semismoothness. Our working assumptions are stated in Section 3. In Section 4 we introduce the generalized differential $\partial_s \Psi$ and investigate some of its properties. In Section 5 a semismoothness and α -order semismoothness concept for the operator Ψ is proposed and studied in detail. The results are illustrated by applications to nonlinear complementarity problems. In particular, we demonstrate the necessity of our assumptions by several (counter-) examples. In Section 6 we propose a Newton-like method for the solution of the nonsmooth operator equation (1.4) and use our semismoothness results to establish its q-superlinear convergence. In the case of a α -order semismooth operator Ψ we prove convergence of q-order $1 + \alpha$. Applications to NCPs are provided as illustrating examples and the computation of smoothing steps is discussed. In Section 7 we show that under appropriate assumptions the composition of semismooth operators is again semismooth and develop two chain rules. Finally, in Section 8, we establish some further properties of our generalized differential.

Notations. Given a Banach space Y , we denote by $\|\cdot\|_Y$ its norm, by B_Y its open unit ball, and by \bar{B}_Y its closed unit ball; in the special case $Y = (\mathbb{R}^n, \|\cdot\|_p)$, we prefer to write B_p^n and \bar{B}_p^n , respectively. On a product space $\prod_i Y_i$, we choose $\|y\|_{\prod_i Y_i} = \sum_i \|y\|_{Y_i}$ as norm. $\mathcal{L}(Y, Z)$ denotes the Banach space of bounded linear operators from the Banach space Y to the Banach space Z , equipped with the operator norm $\|\cdot\|_{Y, Z}$. By $\langle v, w \rangle_\Omega$ we denote the dual pairing between $v \in L^p(\Omega)$ and $w \in L^{p'}(\Omega)$, $1/p + 1/p' = 1$. The indicator function of a measurable set $Q \subset \Omega$, taking the value one on Q and zero on its complement $Q^c = \Omega \setminus Q$, is denoted by $\mathbf{1}_Q$. We write μ for the Lebesgue measure on \mathbb{R}^n . Given a function $w \in L^\infty(\Omega)$ and an operator $A \in \mathcal{L}(Y, L^p(\Omega))$, we define the operator $w \cdot A \in \mathcal{L}(Y, L^p(\Omega))$ that takes $y \in Y$ to the function $\omega \in \Omega \mapsto w(\omega)(Ay)(\omega)$. The Fréchet derivative of an operator H is denoted by H' . For convenience, we will write \sum_i and \prod_i instead of $\sum_{i=1}^m$ and $\prod_{i=1}^m$.

2. Generalized differentials and semismoothness in finite dimensions. We begin with an overview of the semismoothness concept in finite dimensions. Let the vector-valued function $f : \mathbb{R}^k \rightarrow \mathbb{R}^l$ be given. We first collect some notions from nonsmooth analysis. Assume that f is locally Lipschitz continuous. According to Rademacher's theorem, the set $U_f \subset \mathbb{R}^k$ of all points x at which f fails to be differentiable is a Lebesgue null set. Hereby, the fact that f is a mapping between finite-dimensional spaces is crucial. Using this, generalized Jacobians can be constructed:

DEFINITION 2.1. Let f be locally Lipschitz. We define the following generalized Jacobians of f at x :

- (a) The Bouligand (B-) subdifferential:

$$\partial_B f(x) \stackrel{\text{def}}{=} \{M \in \mathbb{R}^{l \times k} : \exists(x_j) \subset \mathbb{R}^k \setminus U_f : x_j \rightarrow x, f'(x_j) \rightarrow M\},$$

where f' denotes the Jacobian of f .

(b) Clarke's generalized Jacobian is the convex hull of $\partial_B f(x)$:

$$\partial f(x) \stackrel{\text{def}}{=} \text{co } \partial_B f(x).$$

(c) Qi's C-subdifferential: $\partial_C f(x) \stackrel{\text{def}}{=} \partial f_1(x) \times \cdots \times \partial f_l(x)$. □

These generalized differentials induce multifunctions $\partial_B f, \partial f, \partial_C f : \mathbb{R}^k \rightrightarrows \mathbb{R}^l$. They have the following properties:

(a) $\partial_B f, \partial f$, and $\partial_C f$ are nonempty-, and compact-valued. Moreover, ∂f and $\partial_C f$ are convex-valued.

(b) The multifunctions $\partial_B f, \partial f$, and $\partial_C f$ are upper semicontinuous. (see Definition 9.2 or [11, p. 29]).

(c) $\partial_B f(x) \subset \partial f(x) \subset \partial_C f(x)$ for all x .

Based on Clarke's generalized Jacobian, Qi [40] and Qi and Sun [42] introduced the following notion of semismoothness:

DEFINITION 2.2. f is *semismooth* at $x \in \mathbb{R}^k$ if it is locally Lipschitz and, for all $h \in \mathbb{R}^k$, the limit

$$\lim_{\substack{M \in \partial f(x+th') \\ h' \rightarrow h, t \rightarrow 0^+}} Mh'$$

exists and is finite. □

The following characterization, however, is more appropriate for our purposes:

PROPOSITION 2.3. *Let f be locally Lipschitz. Then f is semismooth at x if and only if f is directionally differentiable at x and*

$$\max_{M \in \partial f(x+h)} \|f(x+h) - f(x) - Mh\|_2 = o(\|h\|_2) \quad \text{as } h \rightarrow 0.$$

DEFINITION 2.4. f is α -*order semismooth*, $0 < \alpha \leq 1$, at $x \in \mathbb{R}^k$ if it is locally Lipschitz, directionally differentiable at x , and if

$$\max_{M \in \partial f(x+h)} \|f(x+h) - f(x) - Mh\|_2 = O(\|h\|_2^{1+\alpha}) \quad \text{as } h \rightarrow 0.$$

□

It is obvious that useful semismoothness concepts can also be obtained by replacing ∂f by other suitable generalized derivatives. This was investigated in a general framework by Jeyakumar [26, 25]. He introduced the concept of $\partial^* f$ -semismoothness, where $\partial^* f$ is an approximate Jacobian [27]. For the definition of approximate Jacobians we refer to [27]; in the sequel, it is sufficient to know that an approximate Jacobian of $f : \mathbb{R}^k \mapsto \mathbb{R}^l$ is a closed-valued multifunctions $\partial^* f : \mathbb{R}^k \rightrightarrows \mathbb{R}^{l \times k}$ and that $\partial_B f, \partial f$, and $\partial_C f$ are approximate Jacobians.

DEFINITION 2.5. Let $f : \mathbb{R}^k \mapsto \mathbb{R}^l$ be continuous and let be given an approximate Jacobian $\partial^* f$ of f .

(a) The function f is called *weakly $\partial^* f$ -semismooth* at x if

$$(2.1) \quad \sup_{M \in \overline{\text{co}} \partial^* f(x+h)} \|f(x+h) - f(x) - Mh\|_2 = o(\|h\|_2) \quad \text{as } h \rightarrow 0.$$

(b) The function f is *$\partial^* f$ -semismooth* at x if

(i) f is B-differentiable at x (e.g., locally Lipschitz near x and directionally differentiable at x , see [46]), and

(ii) f is weakly $\partial^* f$ -semismooth at x . □

Note that ∂f -semismoothness coincides with semismoothness. Obviously, we can define (weak) $\partial^* f$ -semismoothness of order $1 + \alpha$ by requiring the order $O(\|h\|_2^{1+\alpha})$ in (2.1).

Finally, we consider a Newton-like method for the solution of the nonsmooth equation

$$(2.2) \quad f(x) = 0,$$

where $f : \mathbb{R}^k \rightarrow \mathbb{R}^k$ is weakly $\partial^* f$ -semismooth or weakly $\partial^* f$ -semismooth of the order $1 + \alpha$, respectively, at the solution \bar{x} . For this system of equations, Newton-like methods were developed that converge locally q -superlinearly [25, 39, 40, 42], see also [32, 33]. A representative result is the following.

PROPOSITION 2.6. *Denote by $\bar{x} \in \mathbb{R}^k$ a solution of 2.2 and let the initial point $x_0 \in \mathbb{R}^k$ be given. Consider the following Newton-like iteration:*

For $j = 0, 1, 2, \dots$, choose $M_j \in \partial^ f(x_j)$ and compute $x_{j+1} = x_j + s_j$, where*

$$M_j s_j = -f(x_j).$$

Assume that

- (a) f is weakly $\partial^* f$ -semismooth (or weakly $\partial^* f$ -semismooth of the order $1 + \alpha$) at \bar{x} .
- (b) There exist $\eta > 0$ and $C > 0$ such that, for all $x \in \bar{x} + \eta B_2^k$, every $M \in \partial^* f(x)$ is nonsingular with $\|M^{-1}\|_2 \leq C$ (Regularity assumption).

Then there exists $\delta > 0$ such that for all $x_0 \in \bar{x} + \delta B_2^k$ the above iteration either terminates with $x_j = \bar{x}$ or generates a sequence (x_j) that converges q -superlinearly (or with q -order $1 + \alpha$) to \bar{x} .

REMARK 2.7. In many cases, the approximate Jacobian is upper semicontinuous and compact-valued, in particular if $\partial_B f$, ∂f , or $\partial_C f$ are used. Then it is easy to show that the regularity condition 2.6 (b) is already satisfied if all $M \in \partial^* f(\bar{x})$ are nonsingular. □

3. Assumptions. In the rest of the paper, we will impose the following assumptions on F and ψ :

ASSUMPTION 3.1. There are $1 \leq r \leq r_i < q_i \leq \infty$, $1 \leq i \leq m$, such that

- (a) The operator $F : Y \rightarrow \prod_i L^{r_i}(\Omega)$ is continuously Fréchet differentiable.
- (b) The mapping $y \in Y \mapsto F(y) \in \prod_i L^{q_i}(\Omega)$ is locally Lipschitz continuous, i.e., for all $y \in Y$ there exists an open neighborhood $U = U(y)$ and a constant $L_F = L_F(U)$ such that

$$\sum_i \|F_i(y_1) - F_i(y_2)\|_{L^{q_i}} \leq L_F \|y_1 - y_2\|_Y \quad \text{for all } y_1, y_2 \in U.$$

- (c) The function $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$ is Lipschitz continuous of rank $L_\psi > 0$, i.e.,

$$|\psi(x_1) - \psi(x_2)| \leq L_\psi \|x_1 - x_2\|_1 \quad \text{for all } x_1, x_2 \in \mathbb{R}^m,$$

- (d) ψ is semismooth. □

REMARK 3.2. Since by assumption the domain Ω is bounded, we have the continuous imbedding $L^q(\Omega) \subset L^p(\Omega)$ whenever $1 \leq p \leq q \leq \infty$. □

For semismoothness of order > 0 we will strengthen the Assumptions 3.1 as follows:

ASSUMPTION 3.3. As Assumption 3.1, but with (a) and (d) replaced with:

There exists $\alpha \in (0, 1]$ such that

- (a') The operator $F : Y \rightarrow \prod_i L^{r_i}(\Omega)$ is α -order Hölder continuously Fréchet differentiable.

(d') ψ is α -order semismooth. □

Note that for the special case $Y = \prod_i L^{q_i}(\Omega)$ and $F = \text{id}_Y$ we have

$$\Psi : y \in Y \mapsto \psi(y),$$

and it is easily seen that the Assumptions 3.1 or 3.3, respectively, reduce to (c) and (d) or (c) and (d'), respectively.

Under the Assumptions 3.1, the operator Ψ defined in (1.2) is well defined and locally Lipschitz continuous.

PROPOSITION 3.4. *Let the Assumptions 3.1 hold. Then for all $1 \leq q \leq q_i$, $1 \leq i \leq m$, and thus in particular for $q = r$, the operator Ψ defined in (1.2) maps Y locally Lipschitz continuous into $L^q(\Omega)$.*

Proof. Using Lemma 9.1, we first prove $\Psi(Y) \subset L^q(\Omega)$, which follows from

$$\begin{aligned} \|\Psi(y)\|_{L^q} &= \|\psi(F(y))\|_{L^q} \leq \|\psi(0)\|_{L^q} + \|\psi(F(y)) - \psi(0)\|_{L^q} \\ &\leq c_{q,\infty}(\Omega)|\psi(0)| + L_\psi \sum_i \|F_i(y)\|_{L^q} \\ &\leq c_{q,\infty}(\Omega)|\psi(0)| + L_\psi \sum_i c_{q,q_i}(\Omega) \|F_i(y)\|_{L^{q_i}}. \end{aligned}$$

To establish the local Lipschitz continuity, denote by L_F the local Lipschitz constant in Assumption 3.1 (b) on the set U and let $y_1, y_2 \in U$ be arbitrary. Then, again by Lemma 9.1,

$$\begin{aligned} \|\Psi(y_1) - \Psi(y_2)\|_{L^q} &\leq L_\psi \sum_i \|F_i(y_1) - F_i(y_2)\|_{L^q} \\ &\leq L_\psi \sum_i c_{q,q_i}(\Omega) \|F_i(y_1) - F_i(y_2)\|_{L^{q_i}} \\ &\leq L_\psi L_F \left(\max_{1 \leq i \leq m} c_{q,q_i}(\Omega) \right) \|y_1 - y_2\|_Y. \end{aligned}$$

□

4. An infinite-dimensional generalized differential. For the development of a semismoothness concept for the operator Ψ defined in (1.2) we have to choose an appropriate generalized differential. As we already mentioned in the introduction, our aim is to work with a differential that is as closely connected to finite dimensional generalized Jacobians as possible. Hence, we will propose a generalized differential $\partial_s \Psi$ in such a way that its natural finite-dimensional discretization is Qi's C-subdifferential.

Our construction is motivated by a formal pointwise application of the chain rule. In fact, suppose for the moment that the operator $y \in Y \mapsto F(y) \in C(\bar{\Omega})^m$ be strictly differentiable, where $C(\bar{\Omega})$ denotes the space of continuous functions equipped with the max-norm. Then for fixed $\omega \in \Omega$ the function $f : y \mapsto F(y)(\omega)$ is strictly differentiable with derivative $f'(y) \in \mathcal{L}(Y, \mathbb{R}^m)$,

$$f'(y) : v \mapsto (F'(y)v)(\omega).$$

The chain rule for generalized gradients [11, Thm. 2.3.10] applied to the real-valued mapping $y \mapsto \Psi(y)(\omega) = \psi(f(y))$ yields

$$(4.1) \quad \partial(\Psi(y)(\omega)) \subset \partial\psi(f(y)) \circ f'(y) = \left\{ g \in Y^* \left| \begin{array}{l} \langle g, v \rangle = \sum_i d_i(\omega)(F'_i(y)v)(\omega), \\ d(\omega) \in \partial\psi(F(y)(\omega)) \end{array} \right. \right\}.$$

Furthermore, we can replace ' \subset ' by ' $=$ ' if ψ is regular (e.g., convex or concave) or if the linear operator $f'(y)$ is onto, see [11, Thm. 2.3.10]. Inspired by the idea of the finite-dimensional C-subdifferential, our aim is to define the generalized differential $\partial_s \Psi(y)$ in such a way that for all $M \in \partial_s \Psi(y)$, the linear form $v \mapsto (Mv)(\omega)$ is contained in the set on the right hand side of (4.1).

Following the above motivation, and returning to the general setting of Assumption 3.1, we define the generalized differential $\partial_s \Psi(y)$ in such a way that for all $M \in \partial_s \Psi(y)$, the linear form $v \mapsto (Mv)(\omega)$ is an element of the right hand side in (4.1):

DEFINITION 4.1 (Generalized differential $\partial_s \Psi$). Let the Assumptions 3.1 hold. For Ψ as defined in (1.2) we define the generalized differential $\partial_s \Psi : Y \rightrightarrows \mathcal{L}(Y, L^r)$,

$$(4.2) \quad \partial_s \Psi(y) \stackrel{\text{def}}{=} \left\{ M \in \mathcal{L}(Y, L^r) \left| \begin{array}{l} M : v \mapsto \sum_i d_i \cdot (F'_i(y)v), \\ d \text{ measurable selection of } \partial\psi(F(y)) \end{array} \right. \right\}.$$

□

REMARK 4.2. The subscript 's' is chosen to indicate that this generalized differential is designed for superposition operators. □

The generalized differential $\partial_s \Psi(y)$ is nonempty. To show this, we first prove:

LEMMA 4.3. *Let the Assumption 3.1 (a) hold and let $d \in L^\infty(\Omega)^m$ be arbitrary. Then the operator*

$$M : v \in Y \mapsto \sum_i d_i \cdot (F'_i(y)v)$$

is an element of $\mathcal{L}(Y, L^r)$ and

$$(4.3) \quad \|M\|_{Y, L^r} \leq \sum_i c_{r, r_i}(\Omega) \|d_i\|_{L^\infty} \|F'_i(y)\|_{Y, L^{r_i}}.$$

Proof. By Assumption 3.1 (a) and Lemma 9.1

$$\begin{aligned} \|Mv\|_{L^r} &= \left\| \sum_i d_i \cdot (F'_i(y)v) \right\|_{L^r} \leq \sum_i \|d_i\|_{L^\infty} \|F'_i(y)v\|_{L^r} \\ &\leq \left(\sum_i c_{r, r_i}(\Omega) \|d_i\|_{L^\infty} \|F'_i(y)\|_{Y, L^{r_i}} \right) \|v\|_Y \quad \text{for all } v \in Y, \end{aligned}$$

which shows that (4.3) holds and $M \in \mathcal{L}(Y, L^r)$. □

In a next step, we show that the multifunction

$$\partial\psi(F(y)) : \omega \in \Omega \mapsto \partial\psi(F(y)(\omega)) \subset \mathbb{R}^m$$

is measurable (see Definition 9.3 or [44, p. 160]).

LEMMA 4.4. *Any closed-valued, upper semicontinuous multifunction $\Gamma : \mathbb{R}^k \rightrightarrows \mathbb{R}^l$ is Borel measurable.*

Proof. Let $C \subset \mathbb{R}^l$ be compact. We show that $\Gamma^{-1}(C)$ is closed. To this end, let $x_k \in \Gamma^{-1}(C)$ be arbitrary with $x_k \rightarrow x^*$. Then there exist $z_k \in \Gamma(x_k) \cap C$, and, due to the compactness of C , we achieve by transition to a subsequence that $z_k \rightarrow z^* \in C$. Since $x_k \rightarrow x^*$, upper semicontinuity yields that there exist $\hat{z}_k \in \Gamma(x^*)$ with $(z_k - \hat{z}_k) \rightarrow 0$ and thus $\hat{z}_k \rightarrow z^*$. Therefore, since $\Gamma(x^*)$ is closed, we obtain $z^* \in \Gamma(x^*) \cap C$. Hence, $x^* \in \Gamma^{-1}(C)$, which proves that $\Gamma^{-1}(C)$ is closed and therefore a Borel set. □

COROLLARY 4.5. *The multifunction $\partial\psi(F(y)) : \Omega \rightrightarrows \mathbb{R}$ is measurable.*

Proof. By Lemma 4.4, the compact-valued and upper semicontinuous multifunction $\partial\psi$ is Borel measurable. Now, for all closed sets $C \subset \mathbb{R}^m$, we have, setting $u = F(y) \in \prod_i L^{r_i}(\Omega)$,

$$\partial\psi(F(y))^{-1}(C) = u^{-1}(\partial\psi^{-1}(C)).$$

This set is measurable, since $\partial\psi^{-1}(C)$ is a Borel set and u is a (class of equivalent) measurable function(s). \square

The next result is a direct consequence of Lipschitz continuity, see [11, 2.1.2].

LEMMA 4.6. *Under Assumption 3.1 (c) holds $\partial\psi(x) \subset [-L_\psi, L_\psi]^m$ for all $x \in \mathbb{R}^m$.*

Combining this with Corollary 4.5 yields:

LEMMA 4.7. *Let the Assumptions 3.1 hold. Then for all $y \in Y$, the set*

$$(4.4) \quad K(y) = \{d : \Omega \rightarrow \mathbb{R}^m : d \text{ measurable selection of } \partial\psi(F(y))\}$$

is a nonempty subset of $L_\psi \bar{B}_{L^\infty}^m \subset L^\infty(\Omega)^m$.

Proof. By the Theorem on Measurable Selections [44, Cor. 1C] and Corollary 4.5, $\partial\psi(F(y))$ admits at least one measurable selection $d : \Omega \rightarrow \mathbb{R}^m$, i.e.,

$$d(\omega) \in \partial\psi(F(y)(\omega)) \quad \text{a.e. on } \Omega.$$

From Lemma 4.6 follows $d \in L_\psi \bar{B}_{L^\infty}^m$. \square

We now can prove:

PROPOSITION 4.8. *Under the Assumptions 3.1, for all $y \in Y$ the generalized differential $\partial_s \Psi(y)$ is nonempty and bounded in $\mathcal{L}(Y, L^r)$.*

Proof. Lemma 4.7 ensures that there exist measurable selections d of $\partial\psi(F(y))$ and that all these d are contained in $L_\psi \bar{B}_{L^\infty}^m$. Hence, Lemma 4.3 shows that

$$M : v \mapsto \sum_i d_i \cdot (F'_i(y)v)$$

is in $\mathcal{L}(Y, L^r)$. The boundedness of $\partial_s \Psi(y)$ follows from (4.3). \square

We now have everything at hand to introduce a semismoothness concept that is based on the generalized differential $\partial_s \Psi$. We postpone the investigation of further properties of $\partial_s \Psi$ to the Sections 7 and 8. There, we will establish chain rules, the convex-valuedness, weak compact-valuedness, and the weak graph closedness of $\partial_s \Psi$.

5. Semismoothness in function spaces. In this section, we develop a semismoothness concept for the operator Ψ defined in (1.2). Our notion of semismoothness is similar to Jeyakumar's weak semismoothness in Definition 2.5 (a). In place of the finite-dimensional approximate Jacobian we work with the generalized differential $\partial_s \Psi$. Since we will show in Theorem 8.1 that $\partial_s \Psi$ is convex and closed (even compact) in the weak operator topology, there is no need of taking the closed convex hull of $\partial_s \Psi$ as is done in (2.1).

DEFINITION 5.1. The operator Ψ is *semismooth* at $y \in Y$ if

$$(5.1) \quad \sup_{M \in \partial_s \Psi(y+s)} \|\Psi(y+s) - \Psi(y) - Ms\|_{L^r} = o(\|s\|_Y) \quad \text{as } s \rightarrow 0 \text{ in } Y.$$

Ψ is α -*order semismooth*, $0 < \alpha \leq 1$, at $y \in Y$ if

$$(5.2) \quad \sup_{M \in \partial_s \Psi(y+s)} \|\Psi(y+s) - \Psi(y) - Ms\|_{L^r} = O(\|s\|_Y^{1+\alpha}) \quad \text{as } s \rightarrow 0 \text{ in } Y.$$

\square

This definition is easily extended to general operators between Banach spaces. Of course, an appropriate generalized differential must be available. In this paper, we only deal with the superposition operator Ψ and thus we dispense with a more general definition of semismoothness.

In the following main theorem we establish the semismoothness and the β -order semismoothness, respectively, of the operator Ψ .

THEOREM 5.2.

- (a) *Under the Assumptions 3.1, the operator Ψ is semismooth.*
 (b) *Let the Assumptions 3.3 hold. Assume that there exists $\gamma > 0$ such that the set*

$$\Omega_\varepsilon = \left\{ \omega : \max_{\|h\|_1 \leq \varepsilon} \left(\rho(F(y)(\omega), h) - \varepsilon^{-\alpha} \|h\|_1^{1+\alpha} \right) > 0 \right\}, \quad \varepsilon > 0,$$

with the residual function $\rho : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ given by

$$\rho(x, h) = \max_{z^T \in \partial\psi(x+h)} |\psi(x+h) - \psi(x) - z^T h|,$$

has the following decrease property:

$$(5.3) \quad \mu(\Omega_\varepsilon) = O(\varepsilon^\gamma) \quad \text{as } \varepsilon \rightarrow 0^+,$$

Then the operator Ψ is β -order semismooth with

$$(5.4) \quad \beta = \min \left\{ \frac{\gamma\nu}{1 + \gamma/q_0}, \frac{\alpha\gamma\nu}{\alpha + \gamma\nu} \right\}, \quad \text{where}$$

$$q_0 = \min_{1 \leq i \leq m} q_i, \quad \nu = \frac{q_0 - r}{q_0 r} \quad \text{if } q_0 < \infty, \quad \nu = \frac{1}{r} \quad \text{if } q_0 = \infty.$$

The proof of this theorem will be presented in Section 5.1.

REMARK 5.3. Condition 5.3 requires the measurability of the set Ω_ε , which will be verified in the proof. \square

REMARK 5.4. As we will see in Lemma 5.6, it would be sufficient to require only the β -order Hölder continuity of F' in Assumption 3.3 (a') with $\beta \leq \alpha$ as defined in (5.4). \square

It might be helpful to give an explanation of the abstract condition (5.3) here. For convenient notation, let $x = F(y)(\omega)$. Due to the α -order semismoothness of ψ provided by Assumption 3.3, we have $\rho(x, h) = O(\|h\|_1^{1+\alpha})$ as $h \rightarrow 0$. In essence, Ω_ε is the set of all $\omega \in \Omega$ where there exists $h \in \varepsilon \bar{B}_1^m$ for which this asymptotic behavior is not yet observed, because the remainder term $\rho(x, h)$ exceeds $\|h\|_1^{1+\alpha}$ by a factor of at least $\varepsilon^{-\alpha}$, which grows infinitely as $\varepsilon \rightarrow 0$. From the continuity of the Lebesgue measure it is clear that $\mu(\Omega_\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. The decrease condition (5.3) essentially states that the measure of the set Ω_ε where $F(y)$ takes 'bad values', i.e., values at which the radius of small residual is very small, decreases with the rate ε^γ .

The following Example 5.5 demonstrates the applicability of Theorem 5.2 to nonlinear complementarity problems. It also provides a very concrete interpretation of condition (5.3).

EXAMPLE 5.5 (Application to NCPs). The reformulation of nonlinear complementarity problems (1.5) in the form (1.6) leads to an important special case of the operator equations (1.4) under consideration. Let the operator $Z : L^p(\Omega) \rightarrow L^r(\Omega)$, $1 \leq r < p \leq \infty$, arising in the NCP be given. We choose $Y = L^p(\Omega)$, set $r_1 = r_2 = r$, and define

$$F : y \in Y \mapsto (y, Z(y)) \in L^{r_1}(\Omega) \times L^{r_2}(\Omega).$$

Then (1.6) is equivalent to (1.4) with $\psi \equiv \phi$. Assume that

- (a) The operator $Z : L^p(\Omega) \rightarrow L^r(\Omega)$ is continuously Fréchet differentiable.
- (b) There is $q \in (r, \infty]$ such that $Z : L^p(\Omega) \rightarrow L^q(\Omega)$ is locally Lipschitz continuous.
- (c) ϕ is Lipschitz continuous.
- (d) ϕ is semismooth.

Then the Assumptions 3.1 are satisfied with $q_1 = p$ and $q_2 = q$. In fact, (a) and the continuous embedding $L^p(\Omega) \subset L^r(\Omega)$ imply 3.1 (a). Further, (b) and the Lipschitz continuity of the identity $u \in L^p(\Omega) \mapsto u \in L^p(\Omega)$ yield 3.1(b). Finally, (c),(d) implies 3.1 (c),(d). Therefore, we can apply Theorem 5.2 and obtain that Φ is semismooth:

$$(5.5) \quad \sup_{M \in \partial_s \Phi(y+s)} \|\Phi(y+s) - \Phi(y) - Ms\|_{L^r} = o(\|s\|_{L^p}) \quad \text{as } s \rightarrow 0 \text{ in } L^p(\Omega).$$

Further, we have for all $M \in \partial_s \Phi(u)$ and $v \in Y$

$$(5.6) \quad Mv = d_1 v + d_2 \cdot (Z'(y)v),$$

where $d \in L^\infty(\Omega)^2$ is a measurable selection of $\partial\phi(y, Z(y))$.

In Section 1 we introduced the optimal control problem (1.8) and showed how it can be converted to an equivalent NCP by setting $w = \psi - y$ and choosing $Z(y) = -\nabla J(\psi - y)$. Under mild assumptions on the smoothness of the boundary of Ω given data it was shown in [51] that there exists $p > 2$ such that the unique solution $\bar{y} = \psi - \bar{w}$ of the NCP lies in $L^p(\Omega)$ and that the operator Z satisfies the above assumptions (a), (b) with $q = p$ and $r = 2$.

In the rest of this example we focus on semismoothness of order $\beta > 0$. As above we see that Assumption 3.3 holds if instead of (a) and (d) we require

- (a') The operator $Z : L^p(\Omega) \rightarrow L^r(\Omega)$ is α -Hölder continuously Fréchet differentiable.
- (d') ϕ is α -order semismooth.

If also condition (5.3) is satisfied, we can apply Theorem 5.2 to derive the β -order semismoothness of Φ .

Once we have chosen a particular NCP-function, condition (5.3) can be made very concrete. We discuss this for the Fischer–Burmeister function $\phi = \phi_{FB}$, which is Lipschitz continuous and 1-order semismooth, and thus satisfies Assumptions 3.3 (c) and (d') with $\alpha = 1$. Further, this function is C^∞ on $\mathbb{R}^2 \setminus \{0\}$ with derivatives

$$\nabla\phi(x) = \frac{x}{\|x\|_2} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla^2\phi(x) = \frac{1}{\|x\|_2^3} \begin{pmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_2 & x_1^2 \end{pmatrix}.$$

The eigenvalues of $\nabla^2\phi(x)$ are 0 and $\|x\|_2^{-1}$. In particular, we see that $\|\nabla^2\phi(x)\|_2 = \|x\|_2^{-1}$ explodes as $x \rightarrow 0$. If $0 \notin [x, x+h]$, then Taylor expansion of $\phi(x)$ about $x+h$ yields with appropriate $\tau \in [0, 1]$

$$\rho(x, h) = |\phi(x+h) - \phi(x) - \nabla\phi(x+h)^T h| = \frac{1}{2} |h^T \nabla^2\phi(x+\tau h) h| \leq \frac{\|h\|_2^2}{2 \|x+\tau h\|_2}.$$

Further, $\rho(0, h) = 0$, $\rho(x, 0) = 0$. Our aim is to show that (5.3) is equivalent to the condition

$$(5.7) \quad \mu(\{0 < \|F(y)\|_1 < \varepsilon\}) = O(\varepsilon^\gamma) \quad \text{as } \varepsilon \rightarrow 0.$$

Obviously, this follows easily when we have established the following relation:

$$(5.8) \quad \{0 < \|F(y)\|_1 < \varepsilon\} \subset \Omega_\varepsilon \subset \left\{0 < \|F(y)\|_1 < (1 + 2^{-1/2})\varepsilon\right\}.$$

To show the first inclusion in (5.8), let ω be such that $x = F(y)(\omega)$ satisfies $0 < \|x\|_1 < \varepsilon$ and choose $h = -tx$, where $t \in (1, \sqrt{2})$ is such that $\|h\|_1 \leq \varepsilon$. Then a straightforward calculation yields

$$\rho(x, h) = 2\|x\|_2 \geq \sqrt{2}\|x\|_1 = \frac{\sqrt{2}}{t}\|h\|_1 > \|h\|_1 \geq \varepsilon^{-1}\|h\|_1^2.$$

This implies $\omega \in \Omega_\varepsilon$ and thus proves the first inclusion.

To show the second inclusion in (5.8), let $u = F(y)$. If $u(\omega) = 0$ then certainly $\omega \notin \Omega_\varepsilon$, since then $\rho(u(\omega), \cdot) \equiv 0$. If on the other hand $\|u(\omega)\|_1 \geq (1 + 2^{-1/2})\varepsilon$ then we have for all $h \in \varepsilon\bar{B}_1^2$

$$\rho(u(\omega), h) \leq \frac{\|h\|_2^2}{2\|u(\omega) + \tau h\|_2} \leq \frac{\|h\|_1^2}{\sqrt{2}\|u(\omega) + \tau h\|_1} \leq \varepsilon^{-1}\|h\|_1^2,$$

and thus $\omega \notin \Omega_\varepsilon$.

Having established the equivalence of (5.3) and (5.7), the meaning of (5.3) becomes apparent: The set $\{0 < \|F(y)\|_1 < \varepsilon\}$ on which the decrease rate in measure is assumed is the set of all ω where strict complementarity holds, but is less than ε , i.e., $0 < |y(\omega)| + |Z(y)(\omega)| < \varepsilon$. In a neighborhood of these points the curvature of ϕ is very large since $\|\nabla^2\phi\|$ is big. This requires that $|F(y+s)(\omega) - F(y)(\omega)|$ must be very small in order to have a sufficiently small residual $\rho(F(y)(\omega), F(y+s)(\omega) - F(y)(\omega))$.

We stress that a violation of strict complementarity, i.e., $y(\omega) = Z(y)(\omega) = 0$ does not cause any problems since then $\rho(F(y)(\omega), \cdot) = \rho(0, \cdot) \equiv 0$. \square

5.1. Proof of Theorem 5.2. We can simplify the analysis by exploiting the following fact.

LEMMA 5.6. *Let the Assumptions 3.1 hold and suppose that the operator*

$$\Lambda : u \in \prod_i L^{q_i}(\Omega) \mapsto \psi(u) \in L^r(\Omega)$$

is semismooth. Then the operator $\Psi : Y \rightarrow L^r(\Omega)$ defined in (1.2) is also semismooth. Further, if the Assumptions 3.3 hold and Λ is α -order semismooth then Ψ is α -order semismooth.

Proof. We first observe that, given any $M \in \partial_s \Psi(y+s)$, there is $M_\Lambda \in \partial_s \Lambda(F(y+s))$ such that $M = M_\Lambda F'(y+s)$. In fact, there exists a measurable selection $d \in L^\infty(\Omega)^m$ of $\partial\psi(\omega)$ such that $M = \sum_i d_i \cdot F'_i(y+s)$, and obviously $M_\Lambda : v \mapsto \sum_i d_i v_i$ yields an element of $\partial_s \Lambda(F(y+s))$ with the desired property. A more general chain rule will be established in Theorem 7.2.

Setting $u = F(y)$, $v = F(y+s) - F(y)$, and $w = F(y+s)$, we have

$$\begin{aligned} & \sup_{M \in \partial_s \Psi(y+s)} \|\Psi(y+s) - \Psi(y) - Ms\|_{L^r} \\ & \leq \sup_{M_\Lambda \in \partial_s \Lambda(w)} \|\Lambda(w) - \Lambda(u) - M_\Lambda F'(y+s)s\|_{L^r} \\ & \leq \sup_{M_\Lambda \in \partial_s \Lambda(w)} \|\Lambda(w) - \Lambda(u) - M_\Lambda v\|_{L^r} \\ & \quad + \sup_{M_\Lambda \in \partial_s \Lambda(w)} \|M_\Lambda(F(y+s) - F(y) - F'(y+s)s)\|_{L^r} \stackrel{\text{def}}{=} \rho_\Lambda + \rho_{MF}. \end{aligned}$$

By the local Lipschitz continuity of F and the semismoothness of Λ , we obtain

$$\rho_\Lambda = o(\|v\|_{\prod_i L^{q_i}}) = o(\|s\|_Y) \quad \text{as } s \rightarrow 0 \text{ in } Y.$$

Further, since $d \in L_\psi \bar{B}_{L^\infty}^m$ by Lemma 4.7, we have by Assumption 3.1 (a)

$$\begin{aligned} \|\rho_{MF}\|_{L^r} &\leq L_\psi \sum_i \|F_i(y+s) - F_i(y) - F_i'(y+s)s\|_{L^r} \\ &\leq L_\psi \sum_i c_{r,r_i}(\Omega) \|F_i(y+s) - F_i(y) - F_i'(y+s)s\|_{L^{r_i}} \\ &= o(\|s\|_Y) \quad \text{as } s \rightarrow 0 \text{ in } Y. \end{aligned}$$

This proves the first result.

Now let the Assumptions 3.3 hold and Λ be α -order semismooth. Then ρ_Λ and ρ_{MF} are both of the order $O(\|s\|_Y^{1+\alpha})$, which implies the second assertion. \square

For the proof of Theorem 5.2 we need, as a technical intermediate result, the Borel measurability of the function

$$(5.9) \quad \rho : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}, \quad \rho(x, h) = \max_{z^T \in \partial\psi(x+h)} |\psi(x+h) - \psi(x) - z^T h|.$$

We prove this by showing that ρ is upper semicontinuous. Readers familiar with this type of results might want to skip the proof of Lemma 5.7.

Recall that a function $f : \mathbb{R}^l \rightarrow \mathbb{R}$ is upper semicontinuous at x if

$$\limsup_{x' \rightarrow x} f(x') \leq f(x).$$

Equivalently, f is upper semicontinuous if and only if $\{x : f(x) \geq a\}$ is closed for all $a \in \mathbb{R}$.

LEMMA 5.7. *Let $f : (x, z) \in \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}$ be upper semicontinuous. Moreover, let the multifunction $\Gamma : \mathbb{R}^l \rightrightarrows \mathbb{R}^m$ be upper semicontinuous and compact-valued. Then the function*

$$g : \mathbb{R}^l \rightarrow \mathbb{R}, \quad g(x) = \max_{z \in \Gamma(x)} f(x, z),$$

is well-defined and upper semicontinuous.

Proof. For $x \in \mathbb{R}^l$, let $(z_k) \subset \Gamma(x)$ be such that

$$\lim_{k \rightarrow \infty} f(x, z_k) = \sup_{z \in \Gamma(x)} f(x, z).$$

Since $\Gamma(x)$ is compact, we may assume that $z_k \rightarrow z^*(x) \in \Gamma(x)$. Now, by upper semicontinuity of f ,

$$f(x, z^*(x)) \geq \limsup_{k \rightarrow \infty} f(x, z_k) = \sup_{z \in \Gamma(x)} f(x, z) \geq f(x, z^*(x)).$$

Thus, g is well-defined and there exists $z^* : \mathbb{R}^l \rightarrow \mathbb{R}^m$ with $g(x) = f(x, z^*(x))$.

We now prove the upper semicontinuity of g at x . Let $(x_k) \subset \mathbb{R}^l$ tend to x in such a way that

$$\lim_{k \rightarrow \infty} g(x_k) = \limsup_{x' \rightarrow x} g(x'),$$

and set $z_k = z^*(x_k) \in \Gamma(x_k)$. By the upper semicontinuity of Γ there exists $(\hat{z}_k) \subset \Gamma(x)$ with $(\hat{z}_k - z_k) \rightarrow 0$ as $k \rightarrow \infty$.

Since $\Gamma(x)$ is compact, a subsequence can be selected such that the sequence (\hat{z}_k) , and thus (z_k) , converges to some $\hat{z} \in \Gamma(x)$. Now, using that f is upper semicontinuous and $\hat{z} \in \Gamma(x)$,

$$\limsup_{x' \rightarrow x} g(x') = \lim_{k \rightarrow \infty} g(x_k) = \lim_{k \rightarrow \infty} f(x_k, z_k) = \limsup_{k \rightarrow \infty} f(x_k, z_k) \leq f(x, \hat{z}) \leq g(x).$$

Therefore, g is upper semicontinuous at x . \square

LEMMA 5.8. *Let $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$ be locally Lipschitz continuous. Then the function ρ defined in (5.9) is well-defined and upper semicontinuous.*

Proof. Since $\partial\psi$ is upper semicontinuous and compact-valued, the multifunction

$$(x, h) \in \mathbb{R}^m \times \mathbb{R}^m \mapsto \partial\psi(x + h)$$

is upper semicontinuous and compact-valued as well. Further, the mapping

$$(x, h, z) \mapsto |\psi(x + h) - \psi(x) - z^T h|$$

is continuous, and we may apply Lemma 5.7, which yields the assertion. \square

Proof of Theorem 5.2. By Lemma 5.6, it suffices to prove the semismoothness (of order β) of the operator

$$\Lambda : u \in \prod_i L^{q_i}(\Omega) \mapsto \psi(u) \in L^r(\Omega).$$

(a) Semismoothness: In Lemma 5.8 we showed that the function

$$\rho : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}, \quad \rho(x, h) = \max_{z^T \in \partial\psi(x+h)} |\psi(x + h) - \psi(x) - z^T h|,$$

is upper semicontinuous and thus Borel measurable. Hence, for $u, v \in \prod_i L^{r_i}(\Omega)$, the function $\rho(u, v)$ is measurable. We define the measurable function

$$a = \frac{\rho(u, v)}{\|v\|_1 + \mathbf{1}_{\{v=0\}}}.$$

Since $\rho(u(\omega), v(\omega)) = 0$ whenever $v(\omega) = 0$, we obtain

$$\rho(u, v) = a \|v\|_1.$$

Furthermore,

$$(5.10) \quad a(\omega) = \frac{\rho(u(\omega), v(\omega))}{\|v(\omega)\|_1 + \mathbf{1}_{\{v=0\}}(\omega)} = \frac{o(\|v(\omega)\|_1)}{\|v(\omega)\|_1 + \mathbf{1}_{\{v=0\}}(\omega)} \rightarrow 0 \quad \text{as } v(\omega) \rightarrow 0.$$

Due to the Lipschitz continuity of ψ , we have

$$(5.11) \quad \rho(x, h) \leq 2L_\psi \|h\|_1,$$

which implies $a \in 2L_\psi \bar{B}_{L^\infty}$.

Now let (v_k) tend to zero in the space $\prod_i L^{q_i}(\Omega)$ and set $a_k = a|_{v=v_k}$. Then every subsequence of (v_k) contains itself a subsequence $(v_{k'})$ such that $v_{k'} \rightarrow 0$ a.e. on Ω . By (5.10), this implies $a_{k'} \rightarrow 0$ a.e. on Ω . Since $(a_{k'})$ is bounded in $L^\infty(\Omega)$, we conclude

$$\lim_{k' \rightarrow \infty} \|a_{k'}\|_{L^t} = 0 \quad \text{for all } t \in [1, \infty).$$

Hence, in $L^t(\Omega)$, $1 \leq t < \infty$, zero is an accumulation point of every subsequence of (a_k) . This proves $a_k \rightarrow 0$ in all spaces $L^t(\Omega)$, $1 \leq t < \infty$.

Since the sequence (v_k) , $v_k \rightarrow 0$, was arbitrary, we thus have proven that for all $1 \leq t < \infty$ holds

$$\|a\|_{L^t} \rightarrow 0 \quad \text{as } \|v\|_{\prod_i L^{q_i}} \rightarrow 0.$$

Now we can use Hölder's inequality to obtain

$$(5.12) \quad \begin{aligned} \|\rho(u, v)\|_{L^r(\Omega)} &\leq \sum_i \|av_i\|_{L^r} \leq \sum_i \|a\|_{L^{p_i}} \|v_i\|_{L^{q_i}} \\ &\leq \left(\max_{1 \leq i \leq m} \|a\|_{L^{p_i}} \right) \|v\|_{\Pi_i L^{q_i}} = o(\|v\|_{\Pi_i L^{q_i}}) \quad \text{as } \|v\|_{\Pi_i L^{q_i}} \rightarrow 0, \end{aligned}$$

where $p_i = \frac{q_i r}{q_i - r}$ if $q_i < \infty$ and $p_i = r$ if $q_i = \infty$. Note that here we exploited the fact that $r < q_i$. This proves the semismoothness of Λ .

(b) Semismoothness of order β : We now suppose that the Assumptions 3.3 and, in addition, (5.3) hold. First, note that for fixed $\varepsilon > 0$ the function

$$(x, h) \in \mathbb{R}^m \times \mathbb{R}^m \mapsto \rho(x, h) - \varepsilon^{-\alpha} \|h\|_1^{1+\alpha}$$

is upper semicontinuous and that the multifunction

$$x \in \mathbb{R}^m \mapsto \varepsilon \bar{B}_1^m$$

is compact-valued and upper semicontinuous. Hence, by Lemma 5.7, the function

$$x \in \mathbb{R}^m \mapsto \max_{\|h\|_1 \leq \varepsilon} \left(\rho(x, h) - \varepsilon^{-\alpha} \|h\|_1^{1+\alpha} \right)$$

is upper semicontinuous and therefore Borel measurable. This proves the measurability of the set Ω_ε appearing in (5.3). For $\varepsilon > 0$ and $0 < \beta \leq \alpha$ we define the set

$$\Omega_{\beta\varepsilon} = \left\{ \omega : \rho(u(\omega), v(\omega)) > \varepsilon^{-\beta} \|v(\omega)\|_1^{1+\beta} \right\},$$

and observe that

$$\Omega_{\beta\varepsilon} \subset \Omega_\varepsilon \cup \{\|v\|_1 > \varepsilon\} \stackrel{\text{def}}{=} \Omega_\varepsilon \cup \Omega'_\varepsilon.$$

In fact, let $\omega \in \Omega_{\beta\varepsilon}$ be arbitrary. The nontrivial case is $\|v(\omega)\|_1 \leq \varepsilon$. We then obtain for $h = v(\omega)$

$$\rho(u(\omega), h) > \varepsilon^{-\beta} \|h\|_1^{1+\beta} = \varepsilon^{-\alpha} \varepsilon^{\alpha-\beta} \|h\|_1^{1+\beta} \geq \varepsilon^{-\alpha} \|h\|_1^{\alpha-\beta} \|h\|_1^{1+\beta} = \varepsilon^{-\alpha} \|h\|_1^{1+\alpha},$$

and thus, since $\|h\|_1 \leq \varepsilon$,

$$\max_{\|h\|_1 \leq \varepsilon} \left(\rho(u(\omega), h) - \varepsilon^{-\alpha} \|h\|_1^{1+\alpha} \right) > 0,$$

showing that $\omega \in \Omega_\varepsilon$.

In the case $q_0 = \min_{1 \leq i \leq m} q_i < \infty$ we derive the estimate

$$\begin{aligned} \mu(\Omega'_\varepsilon) &= \mu(\{\|v\|_1 > \varepsilon\}) \leq \|\varepsilon^{-1} \|v\|_1\|_{L^{q_0}(\Omega'_\varepsilon)}^{q_0} \\ &\leq \varepsilon^{-q_0} \left(\max_i c_{q_0, q_i}(\Omega'_\varepsilon) \right)^{q_0} \|v\|_{\Pi_i L^{q_i}}^{q_0} = \varepsilon^{-q_0} O(\|v\|_{\Pi_i L^{q_i}}^{q_0}). \end{aligned}$$

If we choose $\varepsilon = \|v\|_{\Pi_i L^{q_i}}^\lambda$, $0 < \lambda < 1$, then

$$\mu(\Omega_{\beta\varepsilon}) \leq \mu(\Omega_\varepsilon) + \mu(\Omega'_\varepsilon) = O\left(\|v\|_{\Pi_i L^{q_i}}^{2\lambda}\right) + O\left(\|v\|_{\Pi_i L^{q_i}}^{(1-\lambda)q_0}\right).$$

This estimate is also true in the case $q_0 = \infty$ since then $\mu(\Omega'_\varepsilon) = 0$ as soon as $\|v\|_{\Pi_i L^{q_i}} < 1$. This can be seen by noting that then for a.a. $\omega \in \Omega$ holds

$$\|v(\omega)\|_1 \leq \|v\|_1 \leq \|v\|_{\Pi_i L^{q_i}} \leq \|v\|_{\Pi_i L^{q_i}}^\lambda = \varepsilon.$$

Introducing $\nu = \frac{q_0 - r}{q_0 r}$ if $q_0 < \infty$ and $\nu = 1/r$, otherwise, for all $0 < \beta \leq \alpha$, we obtain, using (5.11) and Lemma 9.1

$$\begin{aligned} \|\rho(u, v)\|_{L^r(\Omega_{\beta\varepsilon})} &\leq \|2L_\psi \|v\|_1\|_{L^r(\Omega_{\beta\varepsilon})} \leq 2L_\psi c_{r, q_0}(\Omega_{\beta\varepsilon}) \|v\|_{L^{q_0}(\Omega_{\beta\varepsilon})}^m \\ (5.13) \quad &\leq 2L_\psi \mu(\Omega_{\beta\varepsilon})^\nu \|v\|_{L^{q_0}(\Omega_{\beta\varepsilon})}^m \\ &= O\left(\|v\|_{\Pi_i L^{q_i}}^{1+\gamma\lambda\nu}\right) + O\left(\|v\|_{\Pi_i L^{q_i}}^{1+(1-\lambda)\nu q_0}\right). \end{aligned}$$

Again, we have used here the fact that $r < q_0 \leq q_i$, which allowed us to take advantage of the smallness of the set $\Omega_{\beta\varepsilon}$.

Finally, on $\Omega_{\beta\varepsilon}^c$, $(1 + \beta)r \leq q_0$, $0 < \beta \leq \alpha$, holds with our choice $\varepsilon = \|v\|_{\Pi_i L^{q_i}}^\lambda$

$$\begin{aligned} \|\rho(u, v)\|_{L^r(\Omega_{\beta\varepsilon}^c)} &\leq \left\| \varepsilon^{-\beta} \|v\|_1^{1+\beta} \right\|_{L^r(\Omega_{\beta\varepsilon}^c)} \leq c_{r, \frac{q_0}{1+\beta}}(\Omega_{\beta\varepsilon}^c) \|v\|_{\Pi_i L^{q_i}}^{-\beta\lambda} \|v\|_{L^{q_0}(\Omega_{\beta\varepsilon}^c)}^{1+\beta} \\ &= O\left(\|v\|_{\Pi_i L^{q_i}}^{1+\beta(1-\lambda)}\right). \end{aligned}$$

Therefore,

$$\|\rho(u, v)\|_{L^r} = O\left(\|v\|_{\Pi_i L^{q_i}}^{1+\gamma\lambda\nu}\right) + O\left(\|v\|_{\Pi_i L^{q_i}}^{1+(1-\lambda)\nu q_0}\right) + O\left(\|v\|_{\Pi_i L^{q_i}}^{1+\beta(1-\lambda)}\right).$$

We now choose $0 < \lambda < 1$ and $\beta > 0$ with $\beta \leq \alpha$, $(1 + \beta)r \leq q_0$ in such a way that the order of the right hand side is maximized. In the case $(1 + \alpha)r \geq q_0$ the minimum of all three exponents is maximized for the choice $\beta = \frac{q_0 - r}{r} = \nu q_0$ and $\lambda = \frac{q_0}{\gamma + q_0}$. Then all three exponents are equal to $1 + \frac{\gamma\nu q_0}{\gamma + q_0}$ and thus

$$(5.14) \quad \|\rho(u, v)\|_{L^r} = O\left(\|v\|_{\Pi_i L^{q_i}}^{1 + \frac{\gamma\nu q_0}{\gamma + q_0}}\right).$$

If, on the other hand, $(1 + \alpha)r < q_0$ then the third exponent is smaller than the second one for all $0 < \lambda < 1$ and $0 < \beta \leq \alpha$. Further, it is not difficult to see that under these constraints the first and third exponent become maximal for $\beta = \alpha$ and $\lambda = \frac{\alpha}{\alpha + \gamma\nu}$ and attain the value $1 + \frac{\alpha\gamma\nu}{\alpha + \gamma\nu}$. Hence,

$$(5.15) \quad \|\rho(u, v)\|_{L^r} = O\left(\|v\|_{\Pi_i L^{q_i}}^{1 + \frac{\alpha\gamma\nu}{\alpha + \gamma\nu}}\right).$$

Combining (5.14) and (5.15) proves the β -order semismoothness of Λ with β as in (5.4). \square

5.2. Illustrations. In this section we give two examples to illustrate the above analysis by pointing out the necessity of the main assumptions and by showing that the derived results cannot be improved in several respects.

In order to prevent our examples from being too academical, we will not work with the simplest choices possible. Rather, we will throughout use reformulations of NCPs based on the Fischer–Burmeister function.

The examples address the following items:

- Example 5.9 shows the necessity of the norm gap between L^{q_i} - and L^r -norm.

- Example 5.10 discusses the sharpness of our order of semismoothness β in Theorem 5.2 for varying values of γ .

At the indicated places (5.12) and (5.13) in the above proof we needed the gap between the L^r - and L^{q_i} -norms in order to apply Hölder's inequality. The following example illustrates that Theorem 5.2 does in general not hold if we drop the condition $r_i < q_i$ in the Assumptions 3.1.

EXAMPLE 5.9 (Necessity of the norm gap $r < q_i$). We return to the setting of NCPs as described in Example 5.5. Under the assumptions stated there, we obtain from Theorem 5.2 that the estimate (5.5) holds, where $1 \leq r < q \leq \infty$. Our aim here is to show that the requirement $r < q$ is indispensable in the sense that in general (5.5) is violated for $r \geq q$.

As we will see in Section 6, the estimate (5.5) at a solution y of the NCP is the main tool for proving fast local convergence of Newton's method. Hence, we will construct a simple NCP with a unique solution for which (5.5) fails to hold whenever $r \geq q$. Hereby, we use the Fischer–Burmeister NCP-function ϕ_{FB} defined in (1.7) for the reformulation (1.6) of the NCP.

Let $1 < p \leq \infty$ be arbitrary, choose $\Omega = (0, 1)$, and set

$$F(y)(\omega) = y(\omega) + \omega.$$

Obviously, $\bar{y} \equiv 0$ is the unique solution of the NCP. Choosing $q = p$, $\phi = \phi_{FB}$, and $\alpha = 1$, the Assumptions in Example 1.5—and hence also the Assumption 3.3—are satisfied for all $r \in [1, p)$. To show that the requirement $r < p$ is really necessary to obtain the semismoothness of Φ we will investigate the residual

$$(5.16) \quad R(s)(\omega) \stackrel{\text{def}}{=} \Phi(\bar{y} + s) - \Phi(\bar{y}) - Ms, \quad M \in \partial_s \Phi(\bar{y} + s),$$

at $\bar{y} \equiv 0$ with $s \in L^\infty(\Omega)$, $s \geq 0$, $s \neq 0$. Our aim is to show that for all $r \in [1, \infty]$ holds

$$(5.17) \quad \|R(s)\|_{L^r} = o(\|s\|_{L^p}) \quad \text{as } s \rightarrow 0 \text{ in } L^\infty \quad \implies \quad r < p.$$

A straightforward calculation yields (note that ϕ is smooth on $\mathbb{R}^2 \setminus \{0\}$)

$$|R(s)(\omega)| = \omega - \frac{\omega(s(\omega) + \omega)}{\sqrt{2s(\omega)^2 + 2s(\omega)\omega + \omega^2}}.$$

Now let $0 < \varepsilon < 1$. For the special choice $s_\varepsilon \stackrel{\text{def}}{=} \varepsilon \mathbf{1}_{(0, \varepsilon)}$, i.e., $s_\varepsilon(\omega) = \varepsilon$ for $\omega \in (0, \varepsilon)$ and $s_\varepsilon(\omega) = 0$, otherwise, we obtain

$$\|s_\varepsilon\|_{L^p} = \varepsilon^{\frac{p+1}{p}} \quad (1 < p < \infty), \quad \|s_\varepsilon\|_{L^\infty} = \varepsilon.$$

In particular, $s_\varepsilon \rightarrow 0$ in L^∞ as $\varepsilon \rightarrow 0$. For $0 < \omega < \varepsilon$ holds

$$|R(s_\varepsilon)(\omega)| \geq \omega \left(1 - \sup_{0 < t < 1} \frac{1+t}{\sqrt{2+2t+t^2}} \right) = \frac{5-2\sqrt{5}}{5} \omega \geq \frac{\omega}{10}.$$

Hence, $\|R(s_\varepsilon)\|_{L^\infty} \geq \frac{\varepsilon}{10} \geq \frac{\|s_\varepsilon\|_{L^p}}{10}$, and for all $r \in [p, \infty)$

$$\|R(s_\varepsilon)\|_{L^r} \geq \frac{1}{10} \left(\int_0^\varepsilon \omega^r d\omega \right)^{\frac{1}{r}} = \frac{\varepsilon^{\frac{r+1}{r}}}{10(r+1)^{\frac{1}{r}}} \geq \frac{\|s_\varepsilon\|_{L^p}}{10(r+1)^{\frac{1}{r}}}.$$

Therefore, (5.17) is proven. This shows that in (5.5) the norm on the left must be stronger than on the right. \square

Next, we show that, at least in the case $q_0 \leq (1 + \alpha)r$, the order of our semismoothness result is sharp. By showing this for varying values of γ , we also observe that decreasing values of γ reduce the maximum order of semismoothness exactly as stated in Theorem 5.2. Hence, our result does not overestimate the role of γ .

EXAMPLE 5.10 (Order of semismoothness and its dependence on γ). We consider the following NCP, which generalizes the one in Example 5.9: Let $1 < p \leq \infty$ be arbitrary, set $\Omega = (0, 1)$, and choose

$$F(y)(\omega) = y(\omega) + \omega^\theta, \quad \theta > 0.$$

Obviously, $\bar{y} \equiv 0$ is the unique solution of the NCP. Choosing $q = p$, $\phi = \phi_{FB}$, and $\alpha = 1$, the Assumptions in Example 1.5—and hence also Assumption 3.3—are satisfied for all $r \in [1, p)$.

From $F(\bar{y})(\omega) = (0, \omega^\theta)$ follows that $\gamma = 1/\theta$ is the maximum value for which condition (5.7), and thus the equivalent condition (5.3), is satisfied.

With the residual $R(s)$ as defined in (5.16) we obtain

$$|R(s)(\omega)| = \omega^\theta - \frac{\omega^\theta (s(\omega) + \omega^\theta)}{\sqrt{2s(\omega)^2 + 2s(\omega)\omega^\theta + \omega^{2\theta}}}.$$

For $\varepsilon \in (0, 1)$ and $s_\varepsilon \stackrel{\text{def}}{=} \varepsilon^\theta \mathbf{1}_{(0, \varepsilon)}$ we have

$$\|s_\varepsilon\|_{L^p} = \varepsilon^{\frac{p\theta+1}{p}} \quad (1 < p < \infty), \quad \|s_\varepsilon\|_{L^\infty} = \varepsilon^\theta.$$

Further, for $0 < \omega < \varepsilon$ holds

$$|R(s_\varepsilon)(\omega)| \geq \omega^\theta \left(1 - \sup_{0 < t < 1} \frac{1+t}{\sqrt{2+2t+t^2}} \right) = \frac{5-2\sqrt{5}}{5} \omega^\theta \geq \frac{\omega^\theta}{10}.$$

Hence, for all $r \in [1, p)$

$$\|R(s_\varepsilon)\|_{L^r} \geq \frac{1}{10} \left(\int_0^\varepsilon \omega^{r\theta} d\omega \right)^{\frac{1}{r}} = \frac{\varepsilon^{\frac{r\theta+1}{r}}}{10(r\theta+1)^{\frac{1}{r}}} \geq \frac{\|s_\varepsilon\|_{L^p}^{\frac{pr\theta+p}{pr\theta+r}}}{10(r\theta+1)^{\frac{1}{r}}} = \frac{\|s_\varepsilon\|_{L^p}^{1+\frac{\gamma\nu}{1+\gamma/q_0}}}{10(r\theta+1)^{\frac{1}{r}}}$$

with $q_0 = p = q$, $\gamma = 1/\theta$ and ν as in (5.4). This shows that the value of β given in Theorem 5.2 is sharp for all values of θ (and thus γ) at least as long as $q_0 \leq (1 + \alpha)r$, which in the current setting can be written as $p \leq (1 + \alpha)r$. \square

We think that in the case $q_0 > (1 + \alpha)r$ our value of β could still be slightly improved by splitting Ω in more than the two parts $\Omega_{\beta_\varepsilon}$ and $\Omega_{\beta_\varepsilon}^c$ by choosing different values ε_k for ε that correspond to different powers of $\|v\|_{\Pi_i L^{q_i}}$. In order to keep the analysis as clear as possible, we did not pursue this idea any further in the current paper.

6. Semismooth Newton Method. We now apply the developed semismoothness results to derive a superlinearly convergent Newton-type method for the solution of the nonsmooth operator equation

$$(6.1) \quad \Psi(y) = 0$$

with Ψ as defined in (1.2). Throughout this chapter, let $\bar{y} \in Y$ denote a solution to (6.1). We impose the following regularity condition on $\partial_s \Psi$:

ASSUMPTION 6.1. There exist a Banach space $Y_0 \supset Y$ (Y continuously imbedded) and positive constants $\eta, C_{M^{-1}}$ such that, for all $y \in \bar{y} + \eta B_Y$, every $M \in \partial_s \Psi(u)$ can be extended to an invertible operator $M \in \mathcal{L}(Y_0, L^r)$ with $\|M^{-1}\|_{L^r, Y_0} \leq C_{M^{-1}}$. \square

EXAMPLE 6.2 (Application to NCP). In the following, we want to discuss why the introduction of the additional space Y_0 is of importance. To this end, we consider the reformulation of the NCP (1.5) in the form (1.6) as described in Example 5.5. Recall that the operators $M \in \partial_s \Phi(y)$ assume the form (5.6). Now define $\Omega_1 = \{\omega \in \Omega : d_2(\omega) = 0\}$. Then for almost all $\omega \in \Omega_1$ holds

$$(Mv)(\omega) = d_1(\omega)v(\omega).$$

This shows that (i) M can only be expected to be invertible (between appropriate spaces) if $d_1 \neq 0$ on Ω_1 , and (ii) Mv is in general not more regular (in the L^p -sense) than v and vice versa. Therefore, it is not appropriate to assume that $M \in \mathcal{L}(Y, L^r)$ is invertible as the norm on $Y = L^p$ is stronger than on L^r . However, it is reasonable to assume that M is an L^r -automorphism. This leads to the regularity assumption 6.1 with $Y_0 = L^r(\Omega)$, which can be verified to hold for many NCPs arising in practice, see [51]. In [51] a sufficient condition for regularity is established that is widely applicable and easy to apply. \square

Being aware of the potential gap between Y_0 - and Y -norm, we propose the following Newton method for the solution of (6.1). The algorithm includes a smoothing step to overcome the discrepancy of norms, which will be discussed in Section 6.1.

ALGORITHM 6.3 (Semismooth Newton Method).

0. Choose an initial point $y_0 \in Y$ sufficiently close to a solution $\bar{y} \in Y$ of (6.1). Fix $C_S > 0$ and set $k = 0$.
1. If $\Psi(y_k) = 0$ then stop with solution y_k .
2. Compute $M_k \in \partial_s \Psi(y_k)$, determine $s_k \in Y_0$ by solving

$$M_k s_k = -\Psi(y_k),$$

and set $y_{k+1}^n = y_k + s_k$.

3. Perform a smoothing step:

$$y_{k+1}^n \in Y^0 \mapsto y_{k+1} \in Y \quad \text{such that} \quad \|y_{k+1} - \bar{y}\|_Y \leq C_S \|y_{k+1}^n - \bar{y}\|_{Y_0}.$$

4. Increment k by one and go to Step 1. \square

The local convergence proof for Algorithm 6.3 will clarify the role of the smoothing step.

THEOREM 6.4. *Let the Assumptions 3.1 and 6.1 hold. Then there exists $\delta > 0$ such that for all $y_0 \in \bar{y} + \delta B_Y$ Algorithm 6.3 is well defined and either terminates with a solution y_k of (6.1) or generates a sequence $(y_k) \subset Y$ that converges q -superlinearly to \bar{y} .*

Under the stronger Assumptions 3.3 the rate of convergence is of q -order $1 + \beta$ with $\beta > 0$ given in (5.4).

Proof. Let $y_k \in \bar{y} + \delta B_Y$ with $\delta \in (0, \eta]$ sufficiently small. Then, by Assumption 6.1, the step s_k is well defined. Furthermore, using Assumption 6.1, $\Psi(\bar{y}) = 0$, and Theorem 5.2 gives, as $\delta \rightarrow 0$,

$$\begin{aligned} \|y_{k+1}^n - \bar{y}\|_{Y_0} &= \|y_k - M_k^{-1} \Psi(y_k) - \bar{y}\|_{Y_0} = \|M_k^{-1} (M_k(y_k - \bar{y}) - \Psi(y_k))\|_{Y_0} \\ (6.2) \quad &\leq \|M_k^{-1}\|_{L^r, Y_0} \|0 - \Psi(y_k) - M_k(\bar{y} - y_k)\|_{L^r} \\ &\leq C_{M^{-1}} \|\Psi(\bar{y}) - \Psi(y_k) - M_k(\bar{y} - y_k)\|_{L^r} = o(\|y_k - \bar{y}\|_Y), \end{aligned}$$

and thus, due to the properties of the smoothing step,

$$\|y_{k+1} - \bar{y}\|_Y \leq C_S \|y_{k+1}^n - \bar{y}\|_{Y_0} = o(\|y_k - \bar{y}\|_Y).$$

We conclude: If δ is sufficiently small and $y_0 \in \bar{y} + \delta B_Y$, then inductively, as long as $y_k \neq 0$, the new point y_{k+1} is well defined and $y_{k+1} \in \bar{y} + \delta B_Y$. Furthermore,

$$\|y_{k+1} - \bar{y}\|_Y = o(\|y_k - \bar{y}\|_Y).$$

This establishes the q-superlinear convergence.

Under Assumption 3.3 we can strengthen (6.2) to

$$\|y_{k+1}^n - \bar{y}\|_{Y_0} = O(\|y_k - \bar{y}\|_Y^{1+\beta}) \quad \text{as } k \rightarrow \infty,$$

where β is given by (5.4). Hence, using the properties of the smoothing step,

$$\|y_{k+1} - \bar{y}\|_Y = O(\|y_k - \bar{y}\|_Y^{1+\beta}) \quad \text{as } k \rightarrow \infty,$$

which proves convergence with q-order $1 + \beta$. \square

6.1. Remarks on smoothing steps. The Examples 5.9 and 6.2 demonstrate that the incorporation of a smoothing step into the Newton method can not be avoided. However, since the smoothing step is only needed in pathological cases, it well might happen—and this turns out to be quite common in practice—that these bad situations do not occur very often. Since the design of smoothing steps is by no means trivial and its computation usually requires at least an additional evaluation of F , it would be valuable to have criteria at hand that indicate if a smoothing step is needed or not. The underlying idea is to run the algorithm without smoothing step unless the indicator tells us that a smoothing is required. In the following we discuss several aspects of this issue.

1. If the norms on Y_0 and Y are equivalent, then no smoothing step is needed, i.e., $y_{k+1} = y_{k+1}^n$ can be chosen for all k .
2. If in the k th iteration holds

$$(6.3) \quad \|y_{k+1}^n - \bar{y}\|_Y \leq C_S \|y_{k+1}^n - \bar{y}\|_{Y_0},$$

then the smoothing step can be skipped, i.e., $y_{k+1} = y_{k+1}^n$ can be chosen. However, since \bar{y} is not available, this condition can not be checked at runtime.

3. We now derive a condition that necessarily holds if a smoothing step may be skipped. To this end, assume that y_{k+1}^n satisfies (6.3) and that y_k satisfies the smoothness condition

$$(6.4) \quad \|y_k - \bar{y}\|_Y \leq C_S \|y_k - \bar{y}\|_{Y_0}.$$

Then, as shown in the proof of Theorem 6.4, for any $\kappa > 0$ there is $\delta > 0$ such that for all $y_k \in \bar{y} + \delta B_Y$ holds

$$\begin{aligned} \|y_{k+1}^n - \bar{y}\|_{Y_0} &\leq \kappa \|y_k - \bar{y}\|_Y \leq \kappa C_S \|y_k - \bar{y}\|_{Y_0}, \quad \text{and thus} \\ \|y_{k+1}^n - \bar{y}\|_Y &\leq C_S \|y_{k+1}^n - \bar{y}\|_{Y_0} \leq \kappa C_S \|y_k - \bar{y}\|_Y \leq \kappa C_S^2 \|y_k - \bar{y}\|_{Y_0}. \end{aligned}$$

Therefore,

$$\begin{aligned} \|s_k\|_{Y_0} &\geq \|y_k - \bar{y}\|_{Y_0} - \|y_{k+1}^n - \bar{y}\|_{Y_0} \geq (1 - \kappa C_S) \|y_k - \bar{y}\|_{Y_0}, \\ \|s_k\|_Y &\leq \|y_k - \bar{y}\|_Y + \|y_{k+1}^n - \bar{y}\|_Y \leq (1 + \kappa C_S) C_S \|y_k - \bar{y}\|_{Y_0}, \end{aligned}$$

and for $\kappa < 1/C_S$ we conclude

$$\|s_k\|_Y \leq \frac{1 + \kappa C_S}{1 - \kappa C_S} C_S \|s_k\|_{Y_0} \rightarrow C_S \|s_k\|_{Y_0} \quad \text{as } \kappa \rightarrow 0.$$

We obtain the following result:

LEMMA 6.5. *If, for fixed $\hat{C}_S > C_S$, y_k is sufficiently close to \bar{y} in Y and*

$$(6.5) \quad \|s_k\|_Y > \hat{C}_S \|s_k\|_{Y_0},$$

then at least one of the conditions (6.3), (6.4) is violated. Therefore, if (6.5) occurs and we know from the runtime history of the algorithm that (6.4) is satisfied, we will perform a smoothing step to obtain y_{k+1} from y_{k+1}^n . If, on the other hand, it is not known if y_k satisfies (6.4), we have to return to iteration k and recompute y_k from y_k^n by a smoothing step.

Numerical tests showed that the following simpler rule without backtracking works well in practice: Perform a smoothing step $y_{k+1}^n \mapsto y_{k+1}$ if (6.5) holds and choose $y_{k+1} = y_{k+1}^n$, otherwise.

So far, we did not describe how smoothing steps can be obtained. We do this now for the case of NCP reformulations.

EXAMPLE 6.6 (Smoothing steps for NCPs). We consider operators arising from non-smooth reformulations of NCPs as described in Example 5.5 and further investigated in the Examples 5.9 and 6.2. The following construction of a smoothing step follows an idea in [30], see also [52]. In addition to the assumptions stated in Example 5.5, let us assume that the operator $Z : L^p(\Omega) \rightarrow L^r(\Omega)$ assumes the form $Z(y) = G(y) + \lambda y$, where $\lambda \in L^\infty(\Omega)$ is positive and bounded away from zero, and $G : L^r(\Omega) \mapsto L^p(\Omega)$ is Lipschitz continuous. Note that $G(y)$ is smoother than its preimage y , since $L^p(\Omega) \subset L^r(\Omega)$ with nonequivalent norms. This form of Z arises, e.g., in the first-order necessary optimality conditions of a large class of optimal control problems with bounds on the control and L^2 -regularization [30, 51, 52]. It is well known and easy to verify that $\bar{y} \in L^p(\Omega)$ solves the NCP if and only if

$$S(\bar{y}) \stackrel{\text{def}}{=} (\bar{y} - \lambda^{-1}F(\bar{y}))_+ = \bar{y},$$

where $u_+(\omega) \stackrel{\text{def}}{=} \max\{u(\omega), 0\}$. Further, for all $y \in L^r(\Omega)$ holds $S(y) = \lambda^{-1}G(y)_-$ with $u_- \stackrel{\text{def}}{=} (-u)_+$. Hence, using $|u_- - v_-| \leq |u - v|$, we obtain for all $y \in L^r(\Omega)$

$$|S(y) - \bar{y}| = |S(y) - S(\bar{y})| = \lambda^{-1} |G(y)_- - G(\bar{y})_-| \leq \lambda^{-1} |G(y) - G(\bar{y})|,$$

and therefore

$$\|S(y) - \bar{y}\|_{L^p} \leq \|\lambda^{-1}\|_{L^\infty} \|G(y) - G(\bar{y})\|_{L^p} \leq L_G \|\lambda^{-1}\|_{L^\infty} \|y - \bar{y}\|_{L^r},$$

where L_G is the Lipschitz constant of G . This shows that the mapping $y_k^n \mapsto y_k \stackrel{\text{def}}{=} S(y_k^n)$ is a smoothing step with $C_S = L_G \|\lambda^{-1}\|_{L^\infty}$. \square

7. Semismooth composite operators and chain rules. We consider the scenario where $F = G \circ H$ is a composition of the operators

$$G : X \mapsto \prod_i L^{r_i}(\Omega), \quad H : Y \mapsto X,$$

with X a Banach space, and where $\psi = \psi_1 \circ \psi_2$ is a composition of the functions

$$\psi_1 : \mathbb{R}^l \rightarrow \mathbb{R}, \quad \psi_2 : \mathbb{R}^m \rightarrow \mathbb{R}^l.$$

We impose assumptions on ψ_1, ψ_2, G , and H to ensure that F and ψ satisfy Assumption 3.1. Here is one way to do this:

ASSUMPTION 7.1. There are $1 \leq r \leq r_i < q_i \leq \infty, 1 \leq i \leq m$, such that

- (a) The operators $G : X \rightarrow \prod_i L^{r_i}(\Omega)$ and $H : Y \rightarrow X$ are continuously Fréchet differentiable.
- (b) The operator G maps X locally Lipschitz continuously into $L^{q_i}(\Omega)$.
- (c) The functions ψ_1 and ψ_2 are Lipschitz continuous.
- (d) ψ_1 and ψ_2 are semismooth.

□

It is straightforward to strengthen these assumptions such that they imply the Assumptions 3.3. For brevity, we will not discuss the extension of the next theorem to semismoothness of order β , which is easily established by slight modifications of the assumptions and the proofs.

THEOREM 7.2. *Let the Assumptions 7.1 hold and let $F = G \circ H$ and $\psi = \psi_1 \circ \psi_2$. Then*

- (i) F and ψ satisfy the Assumptions 3.1.
- (ii) Ψ as defined in (1.2) is semismooth.
- (iii) The operator $\Psi_G : z \in X \mapsto \psi(G(z)) \in L^r(\Omega)$ is semismooth and the following chain rule holds:

$$\partial_s \Psi(y) = \partial_s \Psi_G(H(y)) H'(y) = \{M_G H'(y) : M_G \in \partial_s \Psi_G(H(y))\}.$$

- (iv) If $l = 1$ and ψ_1 is strictly differentiable [11, p. 30] then the operator $\Psi_2 : y \in Y \mapsto \psi_2(F(y)) \in L^r(\Omega)$ is semismooth and the following chain rule holds:

$$\partial_s \Psi(y) = \psi_1'(\Psi_2(y)) \partial_s \Psi_2(y) = \{\psi_1'(\Psi_2(y)) \cdot M_2 : M_2 \in \partial_s \Psi_2(y)\}.$$

Proof. (i): 7.1 (a) implies 3.1 (a), 3.1 (b) follows from 7.1 (a),(b), 7.1 (c) implies 3.1 (c), and 3.1 (d) holds by 7.1 (d), since the composition of semismooth functions is semismooth.

(ii): By (i), we can apply Theorem 5.2.

(iii): The Assumptions 7.1 imply the Assumptions 3.1 with G and X instead of F and Y . Hence, Ψ_G is semismooth by Theorem 5.2.

For the proof of the 'c' part of the chain rule, let $M \in \partial_s \Psi(y)$ be arbitrary. By definition, there exists a measurable selection d of $\partial\psi(F(y))$ such that

$$M = \sum_i d_i \cdot F_i'(y).$$

Now, since $F_i'(y) = G_i'(H(y)) H'(y)$,

$$(7.1) \quad \begin{aligned} M &= \sum_i d_i \cdot G_i'(H(y)) H'(y) = M_G H'(y), \quad \text{where} \\ M_G &= \sum_i d_i \cdot G_i'(H(y)). \end{aligned}$$

Obviously, we have $M_G \in \partial_s \Psi_G(H(y))$.

To prove the reverse inclusion, note that any $M_G \in \partial_s \Psi_G(H(y))$ assumes the form (7.1) with appropriate measurable selection $d \in \partial\psi(F(y))$. Then

$$M_G H'(y) = \sum_i d_i \cdot (G_i'(H(y)) H'(y)) = \sum_i d_i \cdot F_i'(y),$$

which shows $M_G H'(y) \in \partial_s \Psi(y)$.

(iv): Certainly, F and ψ_2 satisfy the Assumptions 3.1 (with ψ_2 replaced by ψ). Hence, Theorem 5.2 yields the semismoothness of Ψ_2 . We proceed by noting that a.e. on Ω holds

$$(7.2) \quad \psi_1'(\Psi_2(y)(\omega)) \partial\psi_2(F(y)(\omega)) = \partial\psi(F(y)(\omega)),$$

where we have applied the chain rule for generalized gradients [11, Thm. 2.3.9] and the identity $\partial\psi_1 = \{\psi'_1\}$, see [11, Prop. 2.2.4].

We first prove the ' \supset ' direction of the chain rule. Let $M_2 \in \partial_s \Psi_2$ be arbitrary. It assumes the form

$$M_2 = \sum_i \hat{d}_i \cdot F'_i(y),$$

where $\hat{d} \in L^\infty(\Omega)^m$ is a measurable selection of $\partial\psi_2(F(y))$. Now for any operator M contained in the right hand side of the assertion we have with $d \stackrel{\text{def}}{=} \psi'_1(\Psi_2(y))\hat{d}$

$$M = \psi'_1(\Psi_2(y)) \cdot M_2 = \sum_i d_i \cdot F'_i(y).$$

Obviously, $d \in L^\infty(\Omega)^m$ and, by (7.2), d is a measurable selection of $\partial\psi(F(y))$. Hence, $M \in \partial_s \Psi(y)$.

Conversely, to prove ' \subset ', let $M \in \partial_s \Psi(y)$ be arbitrary and denote by $d \in L^\infty(\Omega)^m$ the corresponding measurable selection of $\partial\psi(F(y))$. Now let $\tilde{d} \in L^\infty(\Omega)^m$ be a measurable selection of $\partial\psi_2(F(y))$ and define $\hat{d} \in L^\infty(\Omega)^m$ by

$$\hat{d}(\omega) = \tilde{d}(\omega) \text{ on } \Omega_0 = \{\omega : \psi'_1(\Psi_2(y)(\omega)) = 0\}, \quad \hat{d}(\omega) = \frac{d(\omega)}{\psi'_1(\Psi_2(y)(\omega))} \text{ on } \Omega \setminus \Omega_0.$$

Then \hat{d} is measurable and $d = \psi'_1(\Psi_2(y))\hat{d}$. Further, $\hat{d}(\omega) = \tilde{d}(\omega) \in \partial\psi_2(F(y))$ on Ω_0 and, using (7.2),

$$\hat{d}(\omega) = \frac{d(\omega)}{\psi'_1(\Psi_2(y)(\omega))} \in \frac{\psi'_1(\Psi_2(y)(\omega))\partial\psi_2(F(y))}{\psi'_1(\Psi_2(y)(\omega))} = \partial\psi_2(F(y)) \text{ on } \Omega \setminus \Omega_0.$$

Thus, \hat{d} is a measurable selection of $\partial\psi_2(F(y))$, and consequently also $\hat{d} \in L^\infty(\Omega)^m$ due to the Lipschitz continuity of ψ_2 . Therefore,

$$M_2 = \sum_i \hat{d}_i \cdot F'_i(y) \in \partial_s \Psi_2(y)$$

and thus $M \in \psi'_1(\Psi_2(y)) \cdot \partial_s \Psi_2(y)$ as asserted. \square

8. Further properties of the generalized differential. We now establish that our generalized differential is convex-valued, weak compact-valued and weakly graph closed. These properties can provide a basis for future research on the connections between $\partial_s \Psi$ and other generalized differentials, in particular the Thibault generalized differential [48] and the Ioffe–Ralph generalized differential [24, 43]. As weak topology on $\mathcal{L}(Y, L^r)$ we use the weak operator topology, which is defined by the seminorms $M \mapsto |\langle w, Mv \rangle_\Omega|$, $v \in Y$, $w \in L^r(\Omega)$, the dual space of $L^r(\Omega)$.

The following result will be of importance.

LEMMA 8.1. *Under Assumption 3.1, the set $K(y)$ defined in (4.4) is convex and weak* sequentially compact in $L^\infty(\Omega)^m$ for all $y \in Y$.*

Proof. From Lemma 4.7 we know that $K(y) \subset L_\psi \bar{B}_L^m$ is nonempty and bounded. Further, the convexity of $\partial\psi(x)$ implies the convexity of $K(y)$. Now let $s_k \in K(y)$ tend to s in $L^2(\Omega)^m$. Then for a subsequence holds $s_{k'}(\omega) \rightarrow s(\omega)$ for a.a. $\omega \in \Omega$. Since $\partial\psi(u(\omega))$ is compact, this implies that for a.a. $\omega \in \Omega$ holds $s(\omega) \in \partial\psi(u(\omega))$ and thus $s \in K(y)$. Hence, $K(y)$ is a bounded, closed, and convex subset of $L^2(\Omega)^m$ and therefore weak sequentially compact in $L^2(\Omega)^m$. Therefore, $K(y)$ is also weak* sequentially closed in

$L^\infty(\Omega)^m$, for, if $(s_k) \subset K(y)$ converges weakly* to s in $L^\infty(\Omega)^m$, then $\langle w, s_k - s \rangle_\Omega \rightarrow 0$ for all $w \in L^1(\Omega)^m \supset L^2(\Omega)^m$, showing that $s_k \rightarrow s$ weakly in $L^2(\Omega)^m$. Thus, $K(y)$ is weak* sequentially closed and bounded in $L^\infty(\Omega)^m$. Since $L^1(\Omega)^m$ is separable, this yields that $K(y)$ is weak* sequentially compact. \square

8.1. Convexity and weak compactness. As further useful properties of $\partial_s \Psi$ we establish the convexity and weak compactness of its images:

THEOREM 8.2. *Under the Assumptions 3.1, the generalized differential $\partial_s \Psi(y)$ is non-empty, convex, and weakly sequentially compact for all $y \in Y$. If Y is separable, then $\partial_s \Psi(y)$ is also weakly compact for all $y \in Y$.*

Proof. The nonemptiness was already stated in Theorem 4.8. The convexity follows immediately from the convexity of the set $K(y)$ derived in Lemma 4.7. We now prove weak sequential compactness. Let $(M_k) \subset \partial_s \Psi(y)$ be any sequence. Then

$$M_k = \sum_i d_{ki} \cdot F'_i(y)$$

with $d_k \in K(y)$, see (4.4). Lemma 8.1 yields that $K(y)$ is weak* sequentially compact in $L^\infty(\Omega)^m$. Hence, we can select a subsequence such that (d_k) converges weak* to $d^* \in K(y)$ in $L^\infty(\Omega)^m$. Define $M^* = \sum_i d_i^* \cdot F'_i(y)$ and observe that $M^* \in \partial_s \Psi(y)$, since $d^* \in K(y)$. It remains to prove that $M_k \rightarrow M^*$ weakly. Let $w \in L^{r'}(\Omega) = L^r(\Omega)'$ and $v \in Y$ be arbitrary. We set $z_i = w \cdot F'_i(y)v$ and note that $z_i \in L^1(\Omega)$. Hence,

$$\begin{aligned} (8.1) \quad |\langle w, (M_k - M^*)v \rangle_\Omega| &\leq \sum_i |\langle w, (d_k - d^*)_i \cdot F'_i(y)v \rangle_\Omega| \\ &= \sum_i |\langle z_i, (d_k - d^*)_i \rangle_\Omega| \longrightarrow 0 \quad \text{as } k \rightarrow \infty. \end{aligned}$$

Therefore, the weak sequential compactness is shown.

By Lemma 4.3, $\partial_s \Psi(y)$ is contained in a closed ball in $\mathcal{L}(Y, L^r)$, on which the weak topology is metrizable if Y is separable (note that $1 \leq r < \infty$ implies that $L^r(\Omega)$ is separable). Hence, in this case the weak compactness follows from the weak sequential compactness. \square

8.2. Weak graph closedness of the generalized differential. Finally, we prove that the multifunction $\partial_s \Psi$ is weakly graph closed:

THEOREM 8.3. *Let the Assumptions 3.1 be satisfied and let $(y_k) \subset Y$ and $(M_k) \subset \mathcal{L}(Y, L^r(\Omega))$ be sequences such that $M_k \in \partial_s \Psi(y_k)$ for all k , $y_k \rightarrow y^*$ in Y , and $M_k \rightarrow M^*$ weakly in $\mathcal{L}(Y, L^r(\Omega))$. Then holds $M^* \in \partial_s \Psi(y^*)$. If, in addition, Y is separable, then the above assertion also holds if we replace the sequences (y_k) and (M_k) by nets.*

Proof. Let $y_k \rightarrow y^*$ in Y and $\partial_s \Psi(y_k) \ni M_k \rightarrow M^*$ weakly. We have the representations $M_k = \sum_i d_{ki} \cdot F'_i(y_k)$ with measurable selections d_k of $\partial \psi(u_k)$, where $u_k = F(y_k)$. We also introduce $u^* = F(y^*)$. The multifunction $\omega \in \Omega \mapsto \partial \psi(u^*(\omega))$ is closed-valued (even compact-valued) and measurable. Furthermore, the function $(\omega, h) \mapsto \|d_k(\omega) - h\|_2$ is a normal integrand on $\Omega \times \mathbb{R}^m$ [44, Cor. 2P]. Hence, by [44, Thm. 2K], the multifunctions $S_k : \Omega \rightarrow \mathbb{R}^m$,

$$S_k(\omega) = \arg \min_{h \in \partial \psi(u^*(\omega))} \|d_k(\omega) - h\|_2$$

are closed-valued (even compact-valued) and measurable. We choose measurable selections s_k of S_k . The sequence (s_k) is contained in the, by Lemma 8.1, sequentially weak* compact set $K(y^*) \subset L^\infty(\Omega)^m$. Further, by Lemma 4.7, we have $d_k \in L_\psi \bar{B}_{L^\infty}^m$.

Hence, by transition to subsequences we achieve $s_k \rightarrow \bar{s} \in K(y^*)$ weak* in $L^\infty(\Omega)^m$ and $d_k \rightarrow \bar{d} \in L_\psi \bar{B}_{L^\infty}^m$ weak* in $L^\infty(\Omega)^m$. Therefore, $(d_k - s_k) \rightarrow (\bar{d} - \bar{s})$ weak* in $L^\infty(\Omega)^m$ and thus also weakly in $L^2(\Omega)^m$. Since $u_k \rightarrow u^*$ in $\prod_i L^{q_i}(\Omega)$, we achieve by transition to a further subsequence that $u_k \rightarrow u^*$ a.e. on Ω . Hence, since $d_k(\omega) \in \partial\psi(u_k(\omega))$ for a.a. $\omega \in \Omega$ and $\partial\psi$ is upper semicontinuous, we obtain from the construction of s_k that $(d_k - s_k) \rightarrow 0$ a.e. on Ω . The sequence $(d_k - s_k)$ is bounded in $L^\infty(\Omega)^m$ and thus the Lebesgue convergence theorem yields $(d_k - s_k) \rightarrow 0$ in $L^2(\Omega)^m$. From $(d_k - s_k) \rightarrow 0$ and $(d_k - s_k) \rightarrow (\bar{d} - \bar{s})$ weakly in $L^2(\Omega)^m$ we see $\bar{d} = \bar{s}$. We thus have

$$d_k \rightarrow \bar{d} = \bar{s} \in K(y^*) \quad \text{weak* in } L^\infty(\Omega)^m.$$

This shows that $\bar{M} \stackrel{\text{def}}{=} \sum_i \bar{d}_i \cdot F'_i(y^*) \in \partial_s \Psi(y^*)$. It remains to prove that $M_k \rightarrow \bar{M}$ weakly. To show this, let $w \in L^{r'}(\Omega) = L^r(\Omega)'$ and $v \in Y$ be arbitrary. Then with $z_{ki} = w \cdot F'_i(y_k)v$ and $z_i = w \cdot F'_i(y^*)v$ holds $z_{ki}, z_i \in L^1(\Omega)$ and

$$\|z_{ki} - z_i\|_{L^1} \leq \|w\|_{L^{r'}} \|F'_i(y_k)v - F'_i(y^*)v\|_{L^r} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Hence, we obtain similar as in (8.1)

$$\begin{aligned} |\langle w, (M_k - \bar{M})v \rangle_\Omega| &\leq \sum_i |\langle w, d_{ki} \cdot F'_i(y_k)v - \bar{d}_i \cdot F'_i(y^*)v \rangle_\Omega| \\ &= \sum_i |\langle d_{ki}, z_{ki} \rangle_\Omega - \langle \bar{d}_i, z_i \rangle_\Omega| \\ &\leq \sum_i (|\langle \bar{d}_i - d_{ki}, z_i \rangle_\Omega| + \|d_{ki}\|_{L^\infty} \|z_i - z_{ki}\|_{L^1}) \rightarrow 0 \quad \text{as } k \rightarrow \infty. \end{aligned}$$

This implies $M^* = \bar{M} \in \partial_s \Psi(y^*)$ and completes the proof of the first assertion.

Now let $(y_\kappa) \subset Y$ and $(M_\kappa) \subset \mathcal{L}(Y, L^r(\Omega))$ be nets such that $M_\kappa \in \partial_s \Psi(y_\kappa)$ for all κ , $y_\kappa \rightarrow y^*$ in Y , and $M_\kappa \rightarrow M$ weakly in $\mathcal{L}(Y, L^r(\Omega))$. Since (y_κ) finally stays in any neighborhood of y^* and since F' is continuous, we see from (4.3) that w.l.o.g. we may assume that (M_κ) is contained in a bounded ball $\mathcal{B} \subset \mathcal{L}(Y, L^r)$. Since, due to the assumed separability of Y , \mathcal{B} is metrizable with respect to the weak topology, we see that we can work with sequences instead of nets. \square

9. Conclusions and future work. In this work, a new semismoothness theory for superposition operators in function spaces was developed. Our semismoothness concept uses a new generalized differential that generalizes Qi's finite-dimensional C-subdifferential. The developed results were shown to be applicable to NCP-function-based reformulations of nonlinear complementarity problems posed in function spaces. Using this semismoothness theory a Newton-like method for nonsmooth operator equations was developed, which, depending on the order of semismoothness of the operator, converges q-superlinearly or with q-order $1 + \alpha$ to a regular solution. Details on the application of this method to the control-constrained optimal control of elliptic PDEs can be found in the author's paper [51]. The numerical results in [51] indicate that the method is efficient and provides mesh-independent performance. We also established the semismoothness of composite operators and developed corresponding chain rules. Furthermore, the multifunction $\partial_s \Psi$ was shown to have several useful properties, in particular weak graph closedness, which can be helpful, e.g., in the development of relationships between $\partial_s \Psi$ and other vector valued generalized differentials.

We plan further investigations in the future. In particular, we will extend the presented Newton method so that it can handle mixed problems of the form

$$\Psi(y) = 0, \quad G(y) = 0,$$

where $G : Y \rightarrow Z$ is a smooth operator. This problem class includes reformulations of Karush–Kuhn–Tucker conditions for many optimal control and variational inequality problems. In our opinion, the main challenge hereby is the choice of a suitable regularity condition on the operators $(M, G'(y))$, $M \in \partial_s \Psi(y)$, and the development of sufficient conditions for regularity that extend the ones given in [50]. Furthermore, it would be interesting to establish the mesh-independence of the proposed semismooth Newton method. We also plan to investigate how our locally convergent Newton method can be made globally convergent in an efficient way. Our preliminary investigations indicate that the merit function $y \in Y \mapsto \|\Psi(y)\|_{L^2}^2 / 2$ is continuously differentiable under reasonable assumptions, which are satisfied, e.g., for $\psi = \phi_{FB}$ and $q_i \geq 2$. Therefore, a convergence theory similar to the one developed in [53] for affine-scaling trust-region methods for bound-constrained nonlinear optimization in function spaces should be transferable to our setting. For the finite-dimensional analogue of the presented algorithm, globalization techniques were developed in, e.g., [14, 16, 28, 50]. Finally, we plan to evaluate the potential of the investigated class of algorithms by extensive numerical tests. Preliminary numerical results for a free boundary value problem and the elliptic control problem (1.8), the latter reported in [51], are very encouraging. The efficient implementation of the algorithm presents further challenges, in particular the fast (approximate) solution of Newton's equation or possible ways of approximating the operators $M_k \in \partial_s \Psi(y_k)$. Depending on the particular problem, multigrid methods can provide a powerful tool for the computation of Newton steps. Alternatively, approximations of M_k could be obtained by approximating the derivatives $F'_i(y_k)$ by quasi-Newton updates.

Appendix.

A consequence of Hölder's inequality. The following estimate is frequently used in our analysis. It follows immediately from Hölder's inequality.

LEMMA 9.1. *Let Ω be bounded, $1 \leq p \leq q \leq \infty$, and*

$$c_{p,q}(\Omega) \stackrel{\text{def}}{=} \mu(\Omega)^{\frac{q-p}{pq}} \text{ if } p < q < \infty, \quad c_{p,\infty}(\Omega) \stackrel{\text{def}}{=} \mu(\Omega)^{1/p} \text{ if } p < \infty, \quad c_{p,q}(\Omega) \stackrel{\text{def}}{=} 1 \text{ if } p = q.$$

Then for all $v \in L^q(\Omega)$ holds

$$\|v\|_{L^p} \leq c_{p,q}(\Omega) \|v\|_{L^q}.$$

Upper semicontinuity and measurability of multifunctions. For convenience, we also provide the definition of upper semicontinuity and measurability of multifunctions [11, 44].

DEFINITION 9.2. A multifunction $\Gamma : U \rightrightarrows \mathbb{R}^l$ defined on $U \subset \mathbb{R}^k$ is *upper semicontinuous* at $x \in U$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\Gamma(x') \subset \{z + h : z \in \Gamma(x), \|h\| < \varepsilon\} \quad \text{for all } x' \in U, \|x' - x\| < \delta.$$

□

DEFINITION 9.3. A multifunction $\Gamma : U \rightrightarrows \mathbb{R}^l$ defined on the measurable set $U \subset \mathbb{R}^k$ is called *measurable* [44, p. 160] if it is closed-valued and if for all closed (or open, or compact, see [44, Prop. 1A]) sets $C \subset \mathbb{R}^l$ the preimage

$$\Gamma^{-1}(C) = \{x \in U : \Gamma(x) \cap C \neq \emptyset\}$$

is measurable.

□

Acknowledgments. This work was done while the author was visiting the Department of Computational and Applied Mathematics and the Center for Research on Parallel Computation at Rice University, which provided an excellent research environment. Especially, the author would like to thank John Dennis and Matthias Heinkenschloss for their hospitality and support.

REFERENCES

- [1] W. ALT, *The Lagrange-Newton method for infinite-dimensional optimization problems*, Numer. Funct. Anal. Optim., 11 (1990), pp. 201–224.
- [2] ———, *Parametric optimization with applications to optimal control and sequential quadratic programming*, Bayreuth. Math. Schr., (1991), pp. 1–37.
- [3] ———, *Sequential quadratic programming in Banach spaces*, in Advances in optimization (Lambrecht, 1991), Springer, Berlin, 1992, pp. 281–301.
- [4] W. ALT AND K. MALANOWSKI, *The Lagrange-Newton method for nonlinear optimal control problems*, Comput. Optim. Appl., 2 (1993), pp. 77–100.
- [5] W. ALT, R. SONTAG, AND F. TRÖLTZSCH, *An SQP method for optimal control of weakly singular Hammerstein integral equations*, Appl. Math. Optim., 33 (1996), pp. 227–252.
- [6] J. APPELL, *The superposition operator in function spaces—a survey*, Exposition. Math., 6 (1988), pp. 209–270.
- [7] J. APPELL AND P. P. ZABREJKO, *Nonlinear superposition operators*, Cambridge University Press, Cambridge, 1990.
- [8] J. M. BORWEIN AND Q. J. ZHU, *A survey of subdifferential calculus with applications*, Nonlinear Anal., 38 (1999), pp. 687–773.
- [9] B. CHEN AND N. XIU, *A global linear and local quadratic noninterior continuation method for nonlinear complementarity problems based on Chen-Mangasarian smoothing functions*, SIAM J. Optim., 9 (1999), pp. 605–623.
- [10] X. CHEN, Z. NASHED, AND L. QI, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, Applied Mathematics Research Report AMR99/7, School of Mathematics, University of New South Wales, Sydney, New South Wales, Australia, 1999.
- [11] F. H. CLARKE, *Optimization and nonsmooth analysis*, John Wiley & Sons Inc., New York, 1983.
- [12] F. H. CLARKE, Y. S. LEDYAEV, R. J. STERN, AND P. R. WOLENSKI, *Nonsmooth analysis and control theory*, Springer-Verlag, New York, 1998.
- [13] B. D. CRAVEN AND B. M. GLOVER, *An approach to vector subdifferentials*, Optimization, 38 (1996), pp. 237–251.
- [14] T. DE LUCA, F. FACCHINEI, AND C. KANZOW, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Programming, 75 (1996), pp. 407–439.
- [15] G. DUVAUT AND J.-L. LIONS, *Inequalities in mechanics and physics*, Springer-Verlag, Berlin, 1976. Grundlehren der Mathematischen Wissenschaften, 219.
- [16] F. FACCHINEI AND C. KANZOW, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Math. Programming, 76 (1997), pp. 493–512.
- [17] F. FACCHINEI AND J. SOARES, *A new merit function for nonlinear complementarity problems and a related algorithm*, SIAM J. Optim., 7 (1997), pp. 225–247.
- [18] M. C. FERRIS AND J.-S. PANG, *Engineering and economic applications of complementarity problems*, SIAM Rev., 39 (1997), pp. 669–713.
- [19] A. FISCHER, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.
- [20] B. M. GLOVER AND D. RALPH, *First order approximations to nonsmooth mappings with application to metric regularity*, Numer. Funct. Anal. Optim., 15 (1994), pp. 599–620.
- [21] R. GLOWINSKI, J.-L. LIONS, AND R. TRÉMOLIÈRES, *Numerical analysis of variational inequalities*, North-Holland Publishing Co., Amsterdam, 1981.
- [22] P. T. HARKER AND J.-S. PANG, *Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications*, Math. Programming, 48 (1990), pp. 161–220.
- [23] M. HEINKENSCHLOSS AND F. TRÖLTZSCH, *Analysis of the Lagrange-SQP-Newton method for the control of a phase field equation*, Control Cybernet., 28 (1999), pp. 177–211.
- [24] A. D. IOFFE, *Nonsmooth analysis: differential calculus of nondifferentiable mappings*, Trans. Amer. Math. Soc., 266 (1981), pp. 1–56.
- [25] V. JEYAKUMAR, *Simple characterizations of superlinear convergence for semismooth equations via approximate Jacobians*, Applied Mathematics Research Report AMR98/28, School of Mathematics, University of New South Wales, Sydney, New South Wales, Australia, 1998.
- [26] ———, *Solving B-differentiable equations*, Applied Mathematics Research Report AMR98/27, School of

- Mathematics, University of New South Wales, Sydney, New South Wales, Australia, 1998.
- [27] V. JEYAKUMAR AND D. T. LUC, *Approximate Jacobian matrices for nonsmooth continuous maps and C^1 -optimization*, SIAM J. Control Optim., 36 (1998), pp. 1815–1832.
- [28] H. JIANG, M. FUKUSHIMA, L. QI, AND D. SUN, *A trust region method for solving generalized complementarity problems*, SIAM J. Optim., 8 (1998), pp. 140–157.
- [29] H. JIANG AND L. QI, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control Optim., 35 (1997), pp. 178–193.
- [30] C. T. KELLEY AND E. W. SACHS, *Multilevel algorithms for constrained compact fixed point problems*, SIAM J. Sci. Comput., 15 (1994), pp. 645–667.
- [31] D. KINDERLEHRER AND G. STAMPACCHIA, *An introduction to variational inequalities and their applications*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980.
- [32] B. KUMMER, *Newton's method for nondifferentiable functions*, in Advances in Mathematical Optimization, J. Guddat et al., eds., Akademie-Verlag, Berlin, 1988, pp. 114–125.
- [33] ———, *Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis*, in Advances in Optimization (Lambrecht, 1991), W. Oettli and D. Pallaschke, eds., Springer, Berlin, 1992, pp. 171–194.
- [34] R. MIFFLIN, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.
- [35] T. S. MUNSON, F. FACCHINEI, M. C. FERRIS, A. FISCHER, AND C. KANZOW, *The Semismooth algorithm for large scale complementarity problems*, Mathematical Programming Technical Report MP-TR-99-07, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1999.
- [36] P. D. PANAGIOTOPOULOS, *Inequality problems in mechanics and applications. Convex and nonconvex energy functions*, Birkhäuser Boston Inc., Boston, Mass., 1985.
- [37] J.-S. PANG, *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Programming, 51 (1991), pp. 101–131.
- [38] ———, *Complementarity problems*, in Handbook of Global Optimization, R. Horst and P. M. Pardalos, eds., Kluwer Acad. Publ., Dordrecht, 1995, pp. 271–338.
- [39] J.-S. PANG AND L. QI, *Nonsmooth equations: motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–465.
- [40] L. QI, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.
- [41] ———, *C-differential operators, C-differentiability and generalized Newton methods*, Research Report AMR96/5, School of Mathematics, University of New South Wales, Sydney, New South Wales, Australia, 1996.
- [42] L. QI AND J. SUN, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.
- [43] D. RALPH, *Rank-1 support functionals and the rank-1 generalized Jacobian, piecewise linear homeomorphisms*, PhD thesis, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1990.
- [44] R. T. ROCKAFELLAR, *Integral functionals, normal integrands and measurable selections*, in Nonlinear Operators and the Calculus of Variations (Summer School, Univ. Libre Bruxelles, Brussels, 1975), J. P. Gossez et al., eds., Springer, Berlin, 1976, pp. 157–207. Lecture Notes in Math., Vol. 543.
- [45] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational analysis*, Springer-Verlag, Berlin, 1998.
- [46] A. SHAPIRO, *On concepts of directional differentiability*, J. Optim. Theory Appl., 66 (1990), pp. 477–487.
- [47] D. SUN AND L. QI, *On NCP-functions*, Comput. Optim. Appl., 13 (1999), pp. 201–220. Computational optimization—a tribute to Olvi Mangasarian, Part II.
- [48] L. THIBAUT, *On generalized differentials and subdifferentials of Lipschitz vector-valued functions*, Nonlinear Anal., 6 (1982), pp. 1037–1053.
- [49] F. TRÖLTZSCH, *An SQP method for the optimal control of a nonlinear heat equation*, Control Cybernet., 23 (1994), pp. 267–288.
- [50] M. ULBRICH, *Non-monotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems*, accepted (in revised form) for publication in SIAM Journal on Optimization.
- [51] ———, *On a nonsmooth Newton method for nonlinear complementarity problems in function space with applications to optimal control*, in Applications and Algorithms of Complementarity, M. C. Ferris, O. L. Mangasarian, and J.-S. Pang, eds., Kluwer Acad. Publ., Dordrecht, (to appear).
- [52] M. ULBRICH AND S. ULBRICH, *Superlinear convergence of affine-scaling interior-point Newton methods for infinite-dimensional nonlinear problems with pointwise bounds*, SIAM J. Control Optim., (to appear).
- [53] M. ULBRICH, S. ULBRICH, AND M. HEINKENSCHLOSS, *Global convergence of trust-region interior-point algorithms for infinite-dimensional nonconvex minimization subject to pointwise bounds*, SIAM J. Control Optim., 37 (1999), pp. 731–764.